Electricity and Magnetism

Melba Phillips, Richard Mara

1966

Preface

This fragmentary and preliminary mate rial fits into an outline of "multi-level monographs" covering those aspects of electromagnetism which in our view an undergraduate physics major should come to know best. The approach is phenomenological and macroscopic, designed to take advantage of prior experience; we begin magnetostatics with magnets, for example. The material is planned on two levels to lead through the four fundamental empirical laws of electricity and magnetism to electromagnetic radiation as a climax. The propagation of electromagnetic disturbances with velocity c, reached in the "first course" material without use of the calculus and equivalent to the homogeneous wave equation, was written in an elementary way by Oliver [1] p. 3, but only recently has appeared in the regular pedagogical literature.

[1] Heaviside, O. (1912). Electromagnetic theory (Vol. 3). Ben

In our treatment we have tried to stress the physical foundations of Maxwell's great synthesis, stating in words the argument corresponding to each mathematical step. This results in a considerably larger proportion of expository writing relative to mathematics than is customarily found in derivations of the wave equation from Maxwell's equations in their usual form. On the other hand, expression of the laws in differential form seems essential for tracing radiation to its sources in a physically meaningful way; the present Chapter 3 of Magnetostatics could be followed almost immediately by Chapter 5 of Monograph III, which would trace radiation fields to retardation effects. We regret having not sufficient time to write such a chapter, as well as the omission of what should have been Chapter 3 of Magnetostatics, an elementary treatment of magnetic materials.

We have assumed no knowledge of special relativity, but have emphasized the necessity for choosing a frame of reference in which to define electric and magnetic field quantities, thus laying a foundation for the historical development of relativity theory. Unlike mechanics, vacuum electrodynamics needs no modification because

OUTLINE OF MONOGRAPHS ON ELECTRICITY AND MAGNETISM

	I. ELECTROSTATICS	II. MAGNETOSTATICS	III. CIRCULATION LAWS AND THEIR CONSEQUENCES
FIRST COURSE MATERIAL	1. Electric Forces and Fields	1. Magnets and Magnetism	1. Faraday's Law of Induction
	2. Electric Energy and Potential	2. Interaction of Steady Currents	2. Ampere's Law Modified
	3. Electrical Properties of Matter	* Magnetic Properties of Matter	3. Propagation of Electromagnetic Disturbances
UPPER DIVISION COURSE MATERIAL	4. Electrostatics Reformulated	3.* Magnetostatics Reformulated	* Maxwell's Equations and Plane Waves, * Radiation Fields

* No textual material was prepared in the summer of 1965 for these chapters.

of special relativity except in interpretation, so that an excursion into relativity theory could be made before or after study of the present material.

The experiments leading to the four fundamental laws are described at some length, but in use this written material should be accompanied by demonstrations and laboratory work. The basic experiments should come to be a part of genuine experience for students, but a laboratory monograph should be written as an extension of the present outline. Ohm's law and circuitry, for example, do not play an appreciable role in any other projected booklets. We cannot overemphasize the importance of laboratory work, although we were not able to undertake detailed consideration of its content.

We assume that students will have studied mechanics, that they know Newton's laws, the definition of work, the meaning of the $\sum$ symbol, and have a working knowledge of elementary vector algebra before our material is introduced. (We do define the vector cross product as if for the first time.) In the material designed for upper-class work we assume basic calculus. All vector calculus is developed as needed, but we attempt throughout to stress the physics, not the mathematics, and attempt no mathematical rigor.

Originally the three parts appear as separate monographs, which are combined in this volume. The three monographs are the three parts in this volume. – *The Mitr*

The first chapters of Monographs I, II, and III should be studied in that order. The few discussion exercises we include can only indicate a type of problem we consider desirable. Numerical problems, which we have made no effort to provide, are also necessary.

M. Phillips
R. T. Mara

Contents

Part I

Electrostatics

Chapter 1

Electric Forces And Fields

In very early times it was observed that certain substances, most strikingly amber after being rubbed with wool, attract chaff and other light objects. What we would call electrostatic effects were particularly troublesome in the spinning of thread spindles were sometimes made of amber, and attracted chaff and dust. According to Pliny, Syrian women called amber itself *harpaga*, "the clutcher," and used the same word for spindle. This was probably the first consistent and repeatable observation of an electrical effect.

But Greek science did not include any study of such odd and usually chance effects, and the science of electricity began with William Gilbert, physician to Queen Elizabeth of England. In his book, *De Magnete* (Figure 1.1), published in 1600, Gilbert carefully distinguished between the behavior of amber and that of a magnet, and showed that the behavior of amber was shared by a great number of substances. It was Gilbert who gave the name *electric* (from the Greek word for amber, *electron*) to the property itself, in order to describe the attraction for light objects shown by glass if rubbed with silk, of wax or resin if rubbed with wool or fur, and so forth for a long list of materials. The title of his book is nevertheless justified: The advances Gilbert was able to make in the knowledge of electricity seem trivial compared with his achievements in disentangling the essential facts of magnetism.

During the great scientific revolution of the seventeenth century, surprisingly little more was learned about electricity. The period was characterized by much writing of a theoretical nature without sufficient recognition of the facts of electrical phenomena, even on the part of such Intellectual giants as Descartes Robert Boyle. The theorizing continued into the eighteenth century, but along with it came the further development of devices for enhancing electrical


GVILIELMI GIL-
BERTI COLCESTREN-
SIS, MEDICI LONDI-
NENSIS,

DE MAGNETE, MAGNETI-
CISQVE CORPORIBVS, ET DE MAG-
no magnete tellure; Phyſiologia noua,
plurimis & argumentis, & expe-
rimentis demonſtrata.

LONDINI
EXCVDEBAT PETRVS SHORT ANNO
MDC.

Figure 1.1: The front folio of *De Magnete* by William Gilbert published in 1600.

effects including various "electrical machines" such as the sphere of Figure 1.2. which can be turned by a handle. The immediate result was that electrical phenomena became an exciting parlor entertainment, but the same devices that shocked (literally!) and delighted ladies and gentlemen in social gatherings also facilitated scientific observations. During the eighteenth century, the essential facts of static electricity became clear.

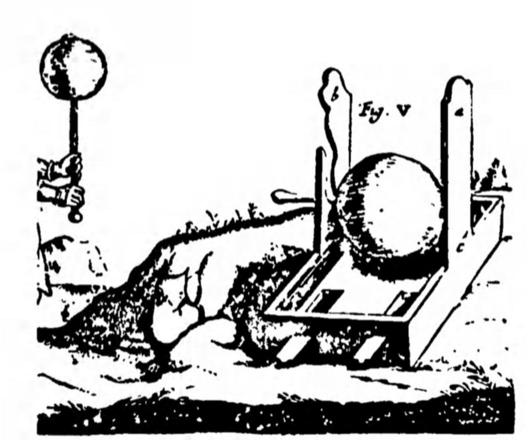


Figure 1.2: A metallic charged sphere can be used for several parlor tricks.

What are the elementary facts? Electrified objects, such as the amber and glass of Gilbert's observations, are said to possess an electric charge, which is of two kinds: Unlike charges attract each other, whereas like charges repel. In naming the two varieties of charge positive (+) and negative (−), Benjamin Franklin was building into the language an important principle: Unlike charges may cancel each other, but the total amount of charge, with due regard for sign, is never changed. If we begin with a piece of matter such as amber or glass which is electrically neutral (exhibits no electrical effects such as attracting fluff or dry chaff), we may electrify it by rubbing with cloth, but the cloth then acquires an equal amount of charge of the opposite sign. (In a problem you are asked to devise a test of this statement.) Charge is neither created nor destroyed, although positive and negative charge can neutralize each other, and the two kinds of charge in a neutral body can often be separated from each other. This is the principle of *charge conservation*, to which no exceptions have ever been found.

Charge is a property of matter which can be described by giving its magnitude and sign, and so is a scalar quantity, like mass, (except that we observe only one kind of mass). Historically, positive charge was defined as that variety which remains on a glass rod if it is rubbed with silk, while negative charge is acquired by amber or sealing wax if rubbed with wool or fur. Franklin found the choice of sign difficult to make, but a choice was necessary to convey the principle of charge conservation. The particular sign convention for charge is actually not important; it *is* important that some sign convention be established and consistently maintained.

Charged bodies exert forces on each other without actual contact, and early in the eighteenth century it was recognized qualitatively that the force between two charged objects decreases as the distance between them is increased. But before quantitative aspects of charge could be investigated, it was necessary to distinguish between *conductors* and *insulators*. Insulators (once called "electrics," now often called "dielectrics") are materials such as amber or glass which can be held in the hand and electrified by rubbing. Conduct-

ors are typically metals, in which charge is free to move and can be conveyed from one part of the material to another. An amber rod is electrified only where it is rubbed, but if a metal sphere on an insulating rod is electrified by stroking it with an electrified amber rod, the charge is distributed over the sphere.

An observation of Franklin's led Joseph Priestley (famous as the discoverer of oxygen) to the first statement of the quantitative relation of electric force to distance. Franklin observed that no electrical effects were to be found *inside* a charged conductor – no forces on a charged pith ball inside a metal can, as indicated in Figure 1.3, for example – except very near the rim. Priestley repeated the experiment with a metal sphere that had a small opening for inserting a test charge such as a charged pith ball suspended on a thread. From the absence of any effect on the charge inside, he concluded that the force between charges varies inversely as the square of the distance between them. In reach ing this conclusion Priestley reasoned by analogy: It was well known that a uniform spherical shell of matter exerts no net gravitational force on a body inside the shell. This result is a geometrical consequence of the in verse square law in three-dimensional space, and the details are left to a problem. The analogy between electric and gravitational forces should be exact, since on symmetry grounds the charge should be distributed uniformly over the surface of a conducting sphere. Priestley's conclusion, published in 1767, is entirely valid, but it received very little attention at the time.

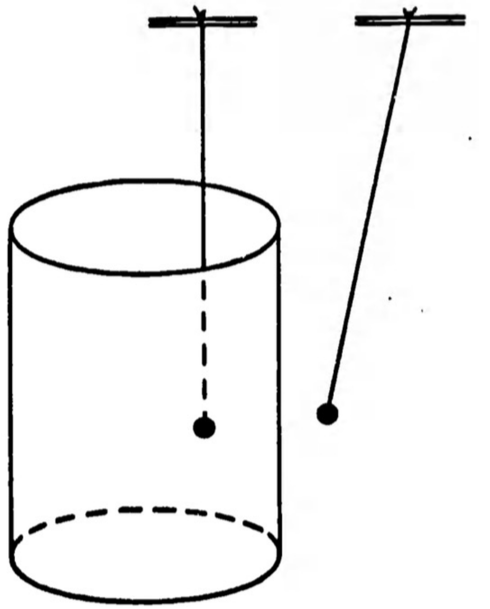


Figure 1.3: A sphere inside a charged metallic can experiences no force.

In 1785, Coulomb measured directly the force between two small charged spheres by means of the torsion balance he had invented (Figure 1.4), and stated the result in terms of quantity of charge as well as distance. For reasons of symmetry two conducting spheres of the same size should share charge equally if they are touched to each other, and an uncharged sphere should take exactly half the charge of an identical sphere when the two are brought into contact. The forces be tween charges of known relative magnitude may then be compared at fixed distance of separation, and Coulomb found that both repulsion and attraction are directly proportional to the product of the two charges involved. The effect of distance is then investigated with two charges of constant magnitude, and the force is found to be inversely proportional to the square of the distance between them, both for repulsion and for attraction.

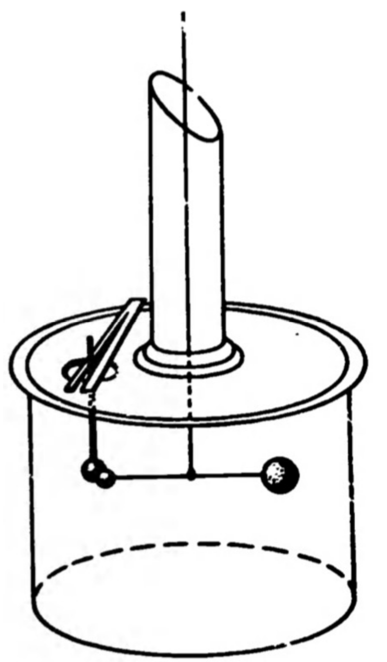


Figure 1.4: A sphere inside a charged metallic can experiences no force.

If q_1 and q_2 represent the magnitude and sign of two fixed and well-localized charges, the magnitude of the force between them is

expressed mathematically as

$$F = \frac{k\, q_1 q_2}{r^2}$$

Here r is the distance from one charge to the other, both taken as points, and the constant k depends on the choice of units for charge, distance, and force. But force is a vector quantity, and the force exerted by q_2 on q_1, is toward or away from q_2 depending on whether the two charges tie opposite in sign or alike. If $\vec{\hat{r}}_{21}$ is a *unit* vector directed from q_2 to q_1 as in Figure 1.5, the force on q_1 at distance r from q_2 is

$$\vec{F} = \frac{k\, q_1 q_2\, \vec{\hat{r}}_{21}}{r^2} \tag{1.1}$$

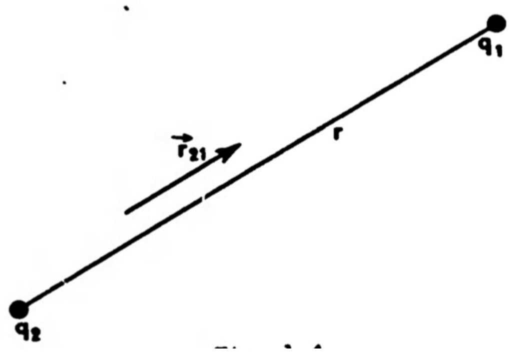


Figure 1.5: The direction of force between two charges is on the unit vector connecting them.

We shall measure q in coulombs, distance in meters, and force in newtons. We shall return to the definition of the coulomb, but may note now that with these units k is found to be 9×10^9, to a very good approximation. It is seen from the equation that the dimensions of k are $\mathrm{N\,m^2/C^2}$, but it will not often be necessary to write out these dimensions if we are consistent in the use of units. The coulomb is clearly a large charge, since two coulombs at a distance of one meter would exert on each ether a force of nearly 10 billion newtons. The stationary charges with which electro*static* experiments are made are small fractions of a coulomb.

If there are more than two charges present, the total force on one of them is found to be the vector sum of the forces exerted by all the others taken separately, as in Figure 1.6. This is equivalent to the statement that the force between two charges is not altered by the presence of a third, and is called the *principle of superposition*. The principle of superposition enables us to find the force on a charge produced by a known fixed distribution of charge, whether well localized or not. The force exerted on q_1 is the vector sum of the effects of all elements Δq of the distribution, each Δq having its own distance r from the position of q_1, and its own unit vector pointed toward that position:

$$\vec{F} = q_1 \sum_{\text{all } \Delta q} \frac{k\, \Delta q\, \vec{\hat{r}}}{r^2} \tag{1.2}$$

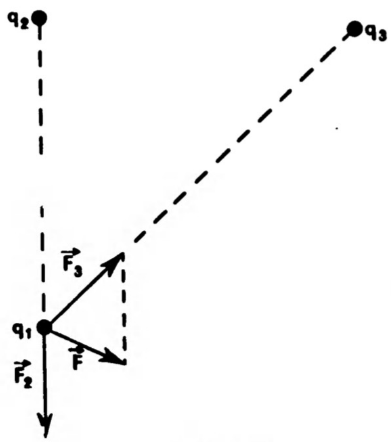


Figure 1.6: Total force on a charge is vector sum of force by other charges.

This force is proportional to q_1, and depends on the position of q_1; for each different position the unit vectors and the distances r from the fixed charges Δq would be different. Yet the size of q_1 may be

considered separately from its position; if $1/2\,q_1$ were substituted for q_1 at any particular place, the force on it would be reduced by one half, but the direction of the force would be the same as before.

The region around a charge or a configuration of charges where electrical effects could be detected has been called a *field*, in much the same way that we might speak of the field of influence of a person, or the field (territory) of a traveling salesman. But the electric field may be made quantitative: Coulomb's law enables us to describe the *field intensity*, as experienced by any small charged body q_1. The field intensity at any point in space is the *force per unit positive charge* placed at that point. It is a vector quantity, designated by $\vec{E}$, which depends only on the fixed distribution of charges producing the field and the position of the point. The field intensity surrounding an isolated point charge q_2 is

$$\vec{E} = \frac{k\,q_2\,\hat{\vec{r}}}{r^2} \tag{1.3}$$

where $\hat{\vec{r}}$ is a unit vector radially outward from the position of q_2 to the point at which $\vec{E}$ is considered (Figure 1.7). If q_2 is a negative charge, the field intensity is directed toward q_2, in the direction of $-\hat{\vec{r}}$.

The field intensity at a particular point P produced by the charges Δq_i of a fixed distribution is

$$\vec{E} = \sum \frac{k\,\Delta q_i\,\hat{\vec{r}}_i}{r_i^2} \tag{1.4}$$

where every unit vector $\hat{\vec{r}}_i$ points from the corresponding element of charge of charge Δq_i toward P, and r_i is the distance from Δq_i to P. Positive elements of charge Δq produce contributions to the total field intensity which are directed away from Δq, whereas contributions from negative source charges are directed toward the sources.

Electric fields are often represented by drawing *field lines*. A field line is a line drawn so that its tan gent is in the direction of the field intensity at each point. The field lines of an isolated point charge are simply straight line radii originating at the position of the point charge (Figure 1.8). The pattern of field lines corresponding to two or more point charges is more .Interesting (Figure 1.9). The number of lines drawn does not matter for our purpose, but the lines begin at positive charges and end at negative charges, and are therefore

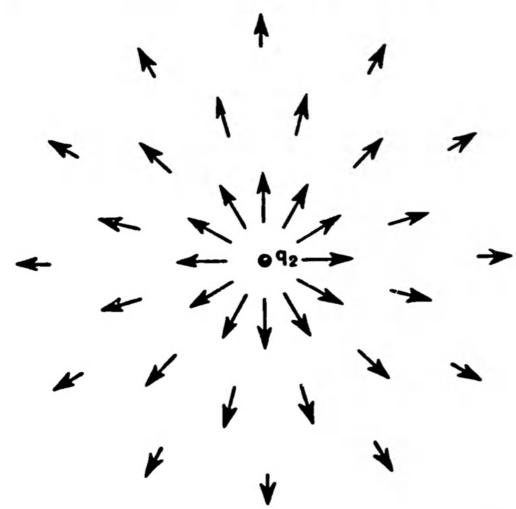


Figure 1.7: Electric field $\vec{E}$ from a charge.

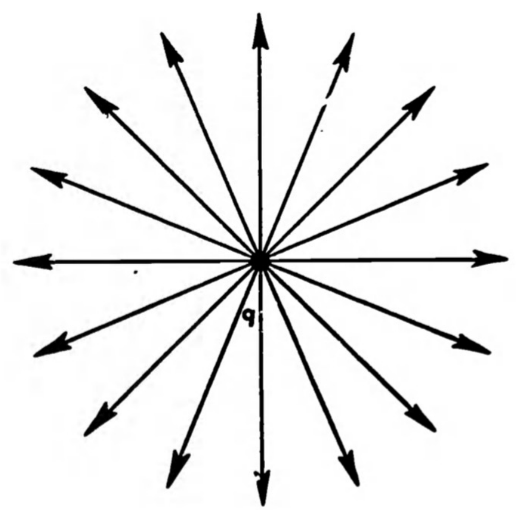


Figure 1.8: Electric field lines of an isolated charge are straight lines.

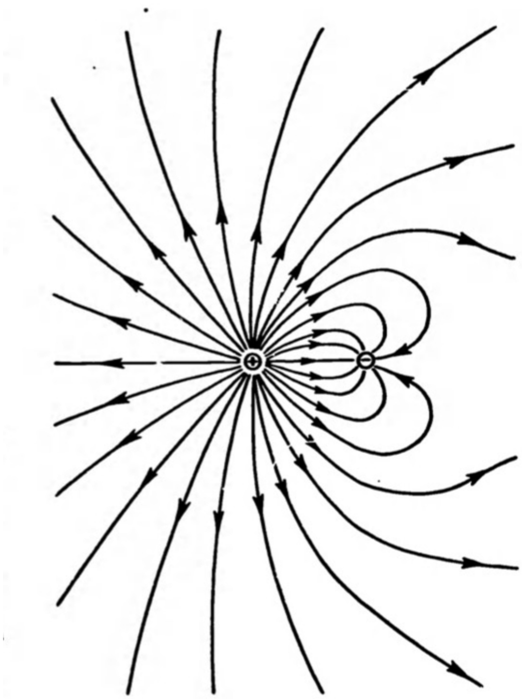


Figure 1.9: Electric field lines of two opposite charges.

closer together near the charges where the field intensity is stronger. In two dimensions we can show only a cross section of the field, which actually extends through three-dimensional space.

There is no particular advantage in introducing the idea of field intensity for applying Coulomb's law to the interaction of two or even several well-localized charges. All we have done is to divide the problem into two problems, the production of a field by a set of charges considered as *sources* of a field intensity $\vec{E}$ at each point, and the force $\vec{F} = q\vec{E}$ experienced by a charge q at some particular point. We shall see almost immediately that there is indeed an advantage in the quantitative definition of $\vec{E}$ for determining the force exerted on a point charge by surface and volume distributions of charge, that hard problems often be come much easier if they are broken up in this way. But the concept of field intensity becomes almost mandatory when we come to consider changing fields, produced by charges which are not stationary, as we shall see in Monograph III of this series.

Let us again consider the electrostatic field intensity whose source is a point charge. The sign and magnitude of the charge is represented by q, and it follows from Coulomb's law that the force per unit positive charge at distance r and in the direction $\vec{\hat{r}}_i$ from q is

$$\vec{E} = \frac{kq\,\vec{\hat{r}}_i}{r^2} \tag{1.3}$$

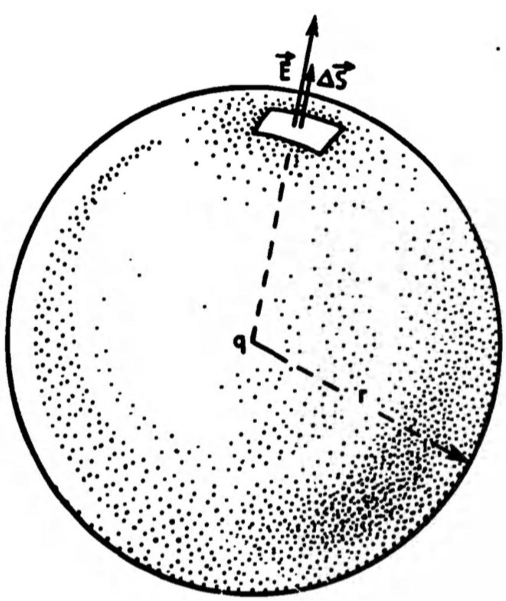


Figure 1.10: Electric field lines of two opposite charges.

A remarkable relation between $\vec{E}$ and its source q can be stated in terms of what is called the *flux* of $\vec{E}$. The meaning of flux can be demonstrated by considering a sphere whose center is at the position of q.

Let the area of the sphere be S, so that a portion of the surface area is ΔS, taken small enough that it is approximately plane. To every such area there corresponds a particular direction, normal to the surface, and thus ΔS may be represented as a vector $\Delta\vec{S}$, whose direction we take positive *outward* from the sphere. The small areas need not be equal in magnitude, but we can specify each area by an index i; the field intensity at the position of ΔS_i is $\vec{E}_i$. We shall call the scalar product $\vec{E}_i \cdot \Delta\vec{S}_i$ the flux of $\vec{E}$ through this particular increment of area, and we can find the total flux of $\vec{E}$ through the surface of the sphere by summing $\vec{E}_i \cdot \Delta\vec{S}_i$ over the entire surface.

If the total flux is called Φ_E, then

$$\Phi_E = \sum_i \vec{E}_i \cdot \Delta\vec{S}_i \tag{1.5}$$

But for our sphere the field intensity $\vec{E}_i$ and $\Delta\vec{S}_i$ are parallel at all points on the surface (Figure 1.10), so that

$$\begin{aligned} \sum_{S \text{ closed}} \vec{E}_i \cdot \Delta\vec{S}_i &= E \times [\text{total area}] \\ &= \frac{kq}{r^2} \cdot 4\pi r^2 \\ &= 4\pi kq \end{aligned}$$

since the magnitude of S at every point on the sphere of radius r is kq/r^2. The total flux through the surface of a sphere is thus independent of the size of the sphere.

But we can carry the idea further to prove that the total flux of S originating from q is the same for *any* closed surface surrounding q. Since the flux through a portion of spherical surface centered at q is independent of the radius, clearly we could make a complicated surface consisting of spherical segments connected by segments of cones without changing the result. Let us now consider a more general surface as shown in Figure 1.11.

The flux through any ΔS is still $\vec{E} \cdot \Delta\vec{S} = E\Delta S\cos\theta$, just the magnitude of $\vec{E}$ times the component of $\Delta\vec{S}$ parallel to $\vec{E}$, or the projection of $\Delta\vec{S}$ on a sphere whose center is q. Thus the shape of the surface does not matter. What does matter is whether the surface surrounds the charge q. If it does, the outward flux of $\vec{E}$ is $4\pi kq$. If the surface surrounds a volume of space outside the charge, the outward flux is equal to the flux into the volume, so that the net flux is zero.

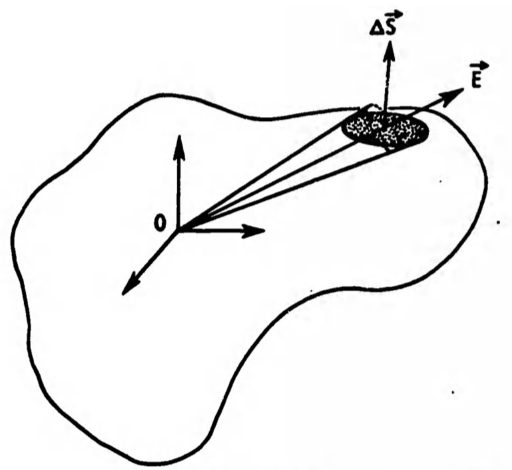


Figure 1.11: Flux through an arbitrary volume containing charge.

That the total flux from a closed surface is just $4\pi kq$ does not depend on the localization of the charge to a particular point or small region. In fact, the superposition principle tells us at once that there may be many point charges, or a charge distribution "smeared out" in space, and that the total flux of $\vec{E}$ through a closed surface is $4\pi k$Q, where Q is the total (net) charge enclosed by the surface. Thus

$$\sum_{S \text{ closed}} \vec{E}_i \cdot \Delta\vec{S}_i = 4\pi k Q \tag{1.6}$$

where $Q = \sum \Delta q$ within the volume en closed. This remarkable result is known as Gauss's law; we note that the physical content

of Gauss's law is the same as Coulomb's law. The original form of the law is essentially a statement of the field intensity in terms of the sources; Gauss's law enables us to locate sources if the field is known. Any closed surface through which there is not net flux contains no net charge, whatever the surface size. But we shall see that Gauss's law also enables us to find very easily the field produced by charge configurations for which the direct sum of the vector field increments is difficult to evaluate.

Before applying Gauss's law to the solution of problems we should note the existence of the geometrical factor 4π. This factor arises from the inverse square law in three-dimensional space: It is simply the area of a sphere divided by r^2. The appearance of 4π is then the consequence of living in a three-dimensional world for any quantity which decreases inversely as the square of the distance from a point source. In this sense the inverse square law is geometrical: The surface density of any quantity which flows from a well-localized (point) source falls off in all directions inversely as the square of the distance from the source, if it is transmitted through space without loss. In the problems you will see that this is true for a stream of particles, and for light.

The geometrical factor 4π in Gauss's theorem will be carried into other equations relating sources and field intensity *unless* the constant k is defined to suppress its appearance. To save writing 4π again and again in these relations the point charge form of Coulomb's law in MKS units is written $k = 1/4\pi\epsilon_0$, so that

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2} \tag{1.7}$$

where $\epsilon_0 = 1/4\pi \times 9 \times 10^9 = 8.85 \times 10^{-12}\ \mathrm{C}^2/(\mathrm{N\,m}^2)$. We note that the factor 4π is written explicitly in one relation to avert its appearance in others. Its appearance somewhere is unavoidable. So far as we are concerned at this stage the coulomb as a unit of charge is a standard arbitrary unit, as is the length of a meter stick. Once all the units have been decided upon, the constant k must be evaluated by experiment, and we have said that in MKS units it is very nearly 9×10^9.

In principle the size of the coulomb could be defined by fixing a value for k in advance, but in practice the unit of charge can be much more accurately defined through the interaction of electric currents, as we shall see in Monograph II.

In MKS units Gauss's law is simply

$$F = \sum_{s \text{ closed}} \vec{E} \cdot \Delta\vec{S} = \frac{Q}{\epsilon_0} \tag{1.8}$$

with Q the total charge enclosed by the surface, as before. Let us apply this theorem to a spherical distribution of charge, for example a charged spherical conductor. Since there is nothing to distinguish one direction of space from another, we can conclude from symmetry considerations that the field intensity at the surface of the mathematical sphere (concentric with the sphere of charge) is directed radially outward, but we do not initially know its magnitude. The total flux of $\vec{E}$ through the surface of a sphere of radius r is therefore simply $4\pi r^2 E$, and by Gauss's law this flux is Q/ϵ_0, where Q is the total charge on the sphere.

But if

$$4\pi r^2 E = \frac{Q}{\epsilon_0}$$

and $\vec{E}$ is directed radially, then

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \frac{Q\vec{\hat{r}}_i}{r^2} \tag{1.9}$$

where $\vec{\hat{r}}_i$ is a unit vector directed radially out from the center of the sphere. This is exactly the same as if Q were a point charge located at the center of the sphere. The same result would be obtained if Q were distributed uniformly, or in any spherically symmetric way, throughout the volume of the sphere. Thus the field of a spherically symmetric distribution of charge is indistinguishable, outside the region of charge, from the field of a point charge (Figure 1.12). This result is by no means obvious from (1.4) with the sum of Δq extending over all regions of the sphere.

The application of Gauss's law to a long cylinder of charge to find the field intensity is again much easier than computation of the sum of field increments from the elements of charge in the cylinder. Consider a cylindrical conductor, for example, so long that its end effects are negligible, and let the Gaussian surface also be a cylinder, its axis coincident with that of the conductor (Figure 1.13). Again we invoke arguments of symmetry: The field intensity $\vec{E}$ is radially symmetric in a plane perpendicular to the axis, and, since the ends of the cylindrical charge are relegated to infinity there is no axial component of $\vec{E}$. There is then no flux through top and bottom

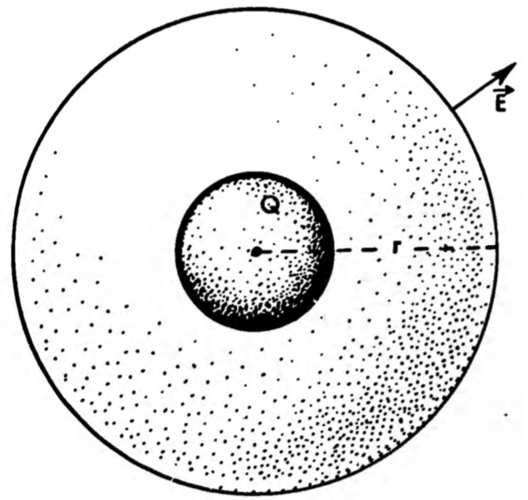


Figure 1.12: Electric field and flux of a charged spherical conductor.

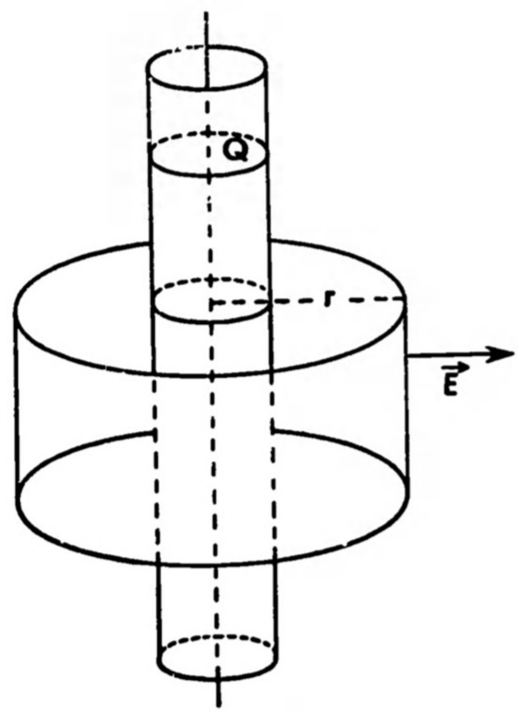


Figure 1.13: Electric field and flux of a charged cylindrical conductor.

or our Gaussian surface, and the magnitude of $\vec{E}$ is the same at all points on the lateral surface. For a Gaussian cylinder of radius r and length L, the total flux is then E times the lateral area, which is $2\pi rL$:

$$\Phi_E = 2\pi rLE = \frac{Q}{\epsilon_0}$$

from which

$$\vec{E} = \frac{1}{2\pi\epsilon_0}\frac{(Q/L)\vec{\hat{r}}_i}{r} \tag{1.10}$$

where $\vec{\hat{r}}_i$ is a unit vector from the center of the cylinder at right angles to the axis, and Q/L is the charge per unit length. Again, as with a sphere, we see that the details of the charge distribution do not matter, so long as the distribution is cylindrically symmetric. Any cylinder of charge produces the same field outside the cylinder as would be produced by an axial line of charge with the same charge per unit length, Q/L. The formula for the field intensity accompanying a line of charge is one we have not previously encountered; at the expense of more trouble the same formula can be derived by summing (1.4).

It may be noted that the fields of both point and line charges become infinite at the source in the limit of geometrical points and lines. Fortunately this need not worry us: Macroscopic charges are always spread out over finite volumes of space, and even elementary charged particles such as the electron are thought to have finite extension in space. And the charge density, defined as the charge per unit volume, must be finite (not infinite) if the total charge of any object is finite. Inside a region of finite charge density the field intensity is also finite, and the determination of the field inside a uniform spherical distribution of charge is left to a problem. The question of the field in side a uniform charge distribution with cylindrical symmetry is equally easy to answer, given Gauss's law.

Let us consider one further application of Gauss's law, a most important one, to find the field just outside the surface of a charged plane conductor. To do so we must examine further the nature of a conductor in electrostatics. A conductor was defined as a substance in which charge is free to move, but in saying *statics* we demand that the charge *not* move. Almost by definition, then, a conductor having a static charge can have no field at all *inside* the conductor, and even at the surface there is no field lying *in* the surface: Any such fields would produce motions of the charge free to move and we would no longer have a static charge. This tells us two things: The net free

charge of a conductor is on its surface, with the interior electrically neutral, and the field intensity just outside the conductor is normal to the conducting surface.

The application of Gauss's law to find the relation of the external field to the surface charge follows immediately. Our Gaussian surface is a short fat cylinder partly inside and partly outside the conductor, its flat external surface parallel to the surface of the conductor (Figure 1.14). Let the cross section area of this cylinder be S. Since the field intensity is normal to S, and all the flux out of the volume passes through this one surface, the total flux is just ES. Therefore,

$$\vec{E} = \frac{(Q/S)\,\vec{n}}{\epsilon_0} \tag{1.11}$$

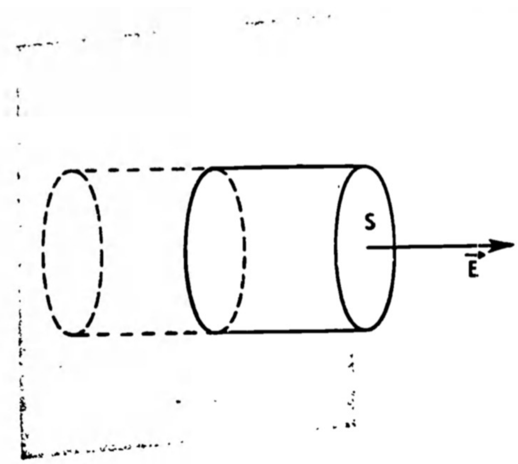


Figure 1.14: Finding relation of external field to the surface charge.

where $\vec{n}$ is a unit vector normal to the surface of the conductor, and Q/S is the charge per unit area of surface. The surface density of charge is such an important quantity in electrostatics that it is often given a special symbol, $\sigma = Q/S$. The magnitude of the field intensity just outside the surface of a conductor is then simply σ/ϵ_0, in terms of the density of surface charge.

If one is close enough to the surface of any conductor, the surface may be considered plane, or, to put it in a different way, if one takes a sufficiently small portion of any surface, the portion may be considered plane. The field just outside the surface is given by (1.11), even for a conducting sphere or cylinder, but then the lines of $\vec{E}$ begin to diverge. For a very large plane, or for a set of two plane conductors with equal and opposite charges, we can obtain a field which is uniform in magnitude and in direction over a considerable volume of space.

Problems

The first nine problems were contributed by A. B. Arons.

1.1 Draw on your own experience to list at least five or six physical situations that you would describe as being associated with "electrification" of various objects. Be sure to identify what was rubbed against what, and indicate what effects provided evidence of the electrification. How can you tell whether or not a particular object is "electrified"? (In addition to cases in which you might have rubbed one material with another, recall circumstances in which you yourself were

involved – scuffling over a carpeted floor, handling dacron or nylon clothes, etc.)

Similarly, list five or six effects that you have heard described as "magnetic"; i.e., describe the behavior of magnets (include some of the things you can and cannot do with them). How can you tell whether or not a particular iron bar is a magnet? What evidence leads us to conclude that we are dealing with different physical effects, justifying the introduction of the two names "electric" and "magnetic"? (Cite some of the *differences* between the two types of phenomena; recall, for example, the unalterable "two endedness" of magnets; the fact that one can hold a magnet in his hand without having it lose its magnetic property, etc.)

1.2 You have undoubtedly heard the word "charge" used many times in connection with electricity. At this point, what meaning do you associate with this term? Can you see "charge" on an object? How can you tell whether or not an object is "charged"? What experiences with electrified bodies lead us to the notion that we might conceive a body as carrying different quantities or "amounts of charge"?

1.3 How do we arrive at the conclusion that "like charges repel"? (What do we mean by "like" charges? How might we set up a situation in which we can assert with confidence that two objects carry like charges? Describe some possible experiments.)

A piece of amber rubbed with wool and a rubber rod rubbed with fur are observed to repel each other. What is the justification for saying that the amber and rubber carry like charges? What do we mean by "unlike" charges? How do we know when two bodies carry unlike charges?

1.4 Describe a hypothetical experimental observation that would force you to say, "Here is a body which carries a third kind of electric charge." (Visualize the interactions between this body and suspended rods of rubber and glass carrying respectively the two kinds of charge we have already recognized.) Under these circumstances what would happen to statements and descriptions based on use of the two adjectives "like" and "unlike"? Outline the nature of the accumulated experience that leads us to believe that only two kinds of electrical charge occur in the physical world. Has this assertion been *proved*?

1.5 During the eighteenth century two other names, "vitreous" and "resinous," competed with Franklin's "positive" and

"negative" for acceptance in the description of electrical phenomena. Why do you think Franklin's terminology finally won the competition? Is there anything wrong with the other terminology?

1.6 Between 1729 and 1736 two English friends, Stephen Gray and Jean Desaguliers, reported the results of a series of experiments "showing that, the Electric Vertue of a glass tube may be conveyed to any other bodies so as to give them the same property of attracting and repelling light bodies as the tube does when excited by rubbing."
They showed, for example, that a cork or other object as much as 800 or 900 feet away could be electrified by connecting it to the glass tube with materials such as metal wires or (moist) hempen string. They found that other materials, such as silk, would not convey the effect. As a matter of fact, they discovered in early painstaking experiments that the distant object would not become electrified if the "transmission line" made contact with the earth and they learned to separate it from the earth by suspending it on silken threads.
Experiences of this kind led investigators to discern that electricity seems to move freely on some materials and not on others. Describe in detail several additional experiments you might per form (or can visualize) with various different objects – experiments that support the findings of Gray and Desaguliers. Why do we introduce the names "conductor" or "non-conductor"? To what experiences do these names refer? Are our own bodies conductors or nonconductors? Cite evidence for your conclusion.

1.7 Suppose we are investigating the force between two small conducting charged spheres A and B of identical size (Figure 1.15). The force is measured by the twist in the suspension fiber when the center of the spheres are 3.00 cm apart. After measuring the force in a given situation when the spheres repel each other and obtaining a value denoted by F_1, we take an identical, but *uncharged*, sphere C on an insulating handle and bring it in contact with sphere B. Then we remove C.
   (a) What are we inclined to say has happened to the quantity of charge carried by sphere B? On what grounds and with what justification?
   (b) In the light of the statements made in the preceding paragraph about Coulomb's investigations, what do we ex-

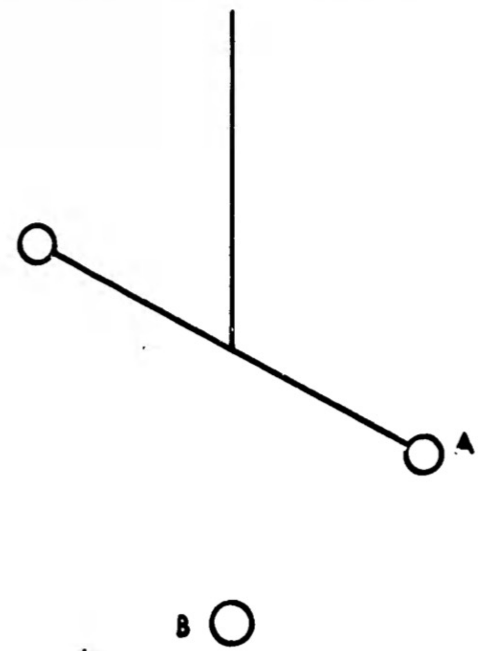


Figure 1.15: Finding force between two small conducting charged spheres.

pect will happen to the magnitude of the force between A and B at the previous center to center separation of 3.00 cm. How will the new value of force compare with F_1 numerically?

(c) Leaving sphere B as it is after this experiment, we discharge C and bring into contact with A. Then remove C as we did before. We measure the force between A and B at 3.00 cm again. How does it now compare with F_1? Describe the results to be expected after additional steps of this kind.

1.8 Suppose we start with a given charged, conducting sphere, A. Now, as we did in Problem 1.7, we bring this sphere in contact with an other, uncharged sphere C, held by an insulating handle. Suppose that C is smaller than A. How do we expect the charge to divide between the two spheres? Suppose C is larger than A? Very much larger than A? (Your responses to these questions are not expected to be numerical. Use words such as "more," "less," "very much less," and explain how you arrive at your inferences.) In the light of this discussion, how would you describe what happens when you touch a charged conducting sphere or establish a conducting path between the sphere and the earth? How would you attempt to describe when transfer of charge from one object to an other ceases? What happens when two differently charged conducting spheres are brought in contact? (Note: These questions do not have simple, pat answers; they will eventually have to be reexamined in a broader context. They are be ing raised at this point to enhance your awareness of some of the problems that lie below the surface of the present discussion.)

1.9 In philosophical discussions of scientific knowledge, it is frequently pointed out that we arrive at the conviction that a particular set of concepts, insights, and descriptions is "correct" not by following one single sequence to a "proof" or an isolated right answer but by finding that the entire network of concepts and experimental observations is internally consistent – that we can criss-cross the network in a variety of different sequences and directions and not develop contradictions. Let's illustrate this notion in connection with our developing conception of electrical phenomena:

Suppose that in a Coulomb torsion balance experiment we charge sphere A on the torsion balance positively. Suppose

that B and C are now observed to exert forces of equal *magnitude* on A at a fixed distance between centers (the forces being, of course, opposite in direction). What would we be led to say about the quantities of unlike charge carried by spheres B and C? If we touch B and C together what would we expect as a final result? It is actually found under such (or analogous) circumstances that the two objects are electrically neutral after contact. In what ways does this reinforce our conceptions of "charge," "quantity of charge," "conservation of charge," neutrality, or unelectrified objects, etc.? The electroscope shown in the diagram (Figure 1.16) consists of two leaves of flexible gold foil suspended on a conducting rod which passes through the insulating stopper of the protective glass flask to a metal cup at the top. It is observed that if an electrified object, such as amber that has been rubbed with wool, is put into the cup without touching, the two gold leaves diverge as indicated, instead of hanging down.

(a) Account for this behavior.
(b) What will happen if the amber rod is removed, without having touched the metal cup?
(c) Could you use this apparatus to test whether the charge acquired by silk used in electrify ing a glass rod is equal and opposite to that acquired by the glass rod? How?

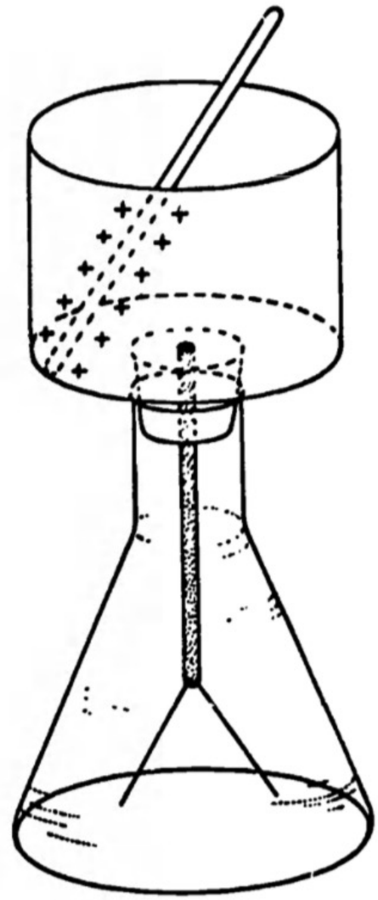


Figure 1.16: An electroscope.

1.10 Show that a uniform spherical shell on matter exerts no net gravitational force on a point mass m inside by considering a double cone of very small aperture, its apex at the point mass, which cuts through the sphere on either side Figure 1.17). What is the total gravitational force on the point mass due to the portions of the spherical shell inside the double cone? Complete the argument to include all the mass in the shell.

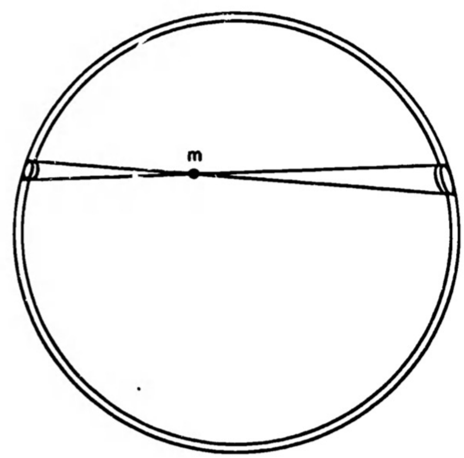


Figure 1.17: An uniform spherical shell.

1.11 (a) Show by applying Gauss's law to the interior of a hollow, conducting, charged sphere that the field inside the sphere is zero.
(b) Assume that you have a uniform volume distribution through out a sphere. (This is not a conductor!) Show that the field in tensity inside the sphere is given by

$$\vec{E} = \frac{r\rho\vec{\hat{r}}}{3\epsilon_0}$$

where $\vec{\hat{r}}$ is a radial unit vector and ρ is the charge density, or charge per unit volume. (Simply apply Gauss's law

to a spherical surface of smaller radius than the radius of the charge distribution.)

1.12 (a) Use Gauss's law to show that the field intensity inside a uniform cylindrical shell of charge is zero.

(b) Assume you have a uniform volume distribution of charge throughout a long circular cylinder, and show that $\vec{E}$ inside the cylinder (not near the ends) is

$$\vec{E} = \frac{r\rho\vec{\hat{r}}}{2\epsilon_0}$$

where $\vec{\hat{r}}$ is an unit vector radially directed out from the axis of the cylinder, and ρ is again the charge density.

1.13 If light is emitted constantly and uniformly in all directions by a spherical source, and there is no intervening material to absorb or reflect it, show that the intensity (amount of light per unit area received by a surface at right angles to its direction) falls off inversely as the square of the distance from the center of the source.

Chapter 2

Electric Energy And Potential

The science of electrostatics began with the study of small and often chance effects such as the attraction of amber for thread and chaff, but in the early days of our planet electrical energy may have played an essential role in the beginning of life Itself. According to one theory, incessant flashes of lightning through an atmosphere of nitrogen, carbon dioxide, and water vapor produced the first organic molecules, from which organized life developed. The frequency and intensity of lightning have diminished, but. a lightning flash today obviously re leases an enormous amount of heat and light. The connection between this impressive natural phenomenon and the behavior of amber was established in the eighteenth century by Benjamin Franklin. The weather is a complicated matter indeed, which scientists are only beginning to understand, but the recognition that lightning is electrical certainly extends electrical phenomena beyond the small theater of parlor entertainment.

In this chapter we are concerned with one aspect of electric energy: work done by, or against, electrostatic forces. That those forces are really very strong is usually masked by the charge neutrality of most objects. Equal amounts of positive and negative charge, however great the amounts separately, can *neutralize* the effects of each other if they are nearly coincident; this is equivalent to the statement that lines of electric field intensity end, as well as begin, and that ordinarily there is just about as much negative charge as positive charge in any particular region. But if there is the possibility of strong forces there is also the possibility of utilizing these forces to do work. In addition to any possible utility, we shall see that

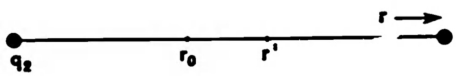


Figure 2.1: An uniform spherical shell.

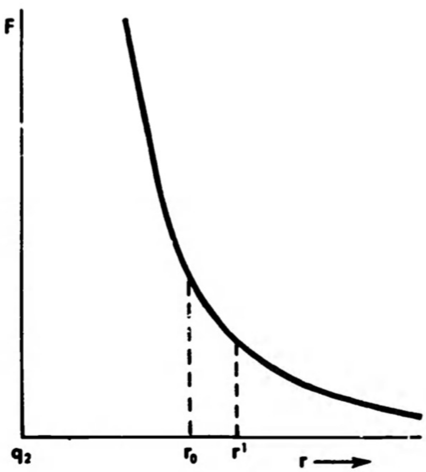


Figure 2.2: Coulomb law force as a function of distance between the charges.

in many instances it is simpler to describe electrical phenomena in terms of work and energy than in terms of forces. For one thing, energy is a scalar quantity while force is a vector, requiring three numbers to specify it instead of only one.

Let us consider the energy associated with the interaction of two point charges. Work is performed against the force of repulsion in bringing two like charges closer to each other, that is, in decreasing their separation r. How much work? We cannot get the answer by simple multiplication of the force given by Coulomb's law and the relative displacement, since the force itself depends on the distance between them and is not the same at the beginning of the displacement as at the end. Rigorous derivation of the formula for computing the energy expended to displace the charge q_1 from r' to r_0 as indicated in Figure 2.1 requires use of the calculus. Here we shall assume the correct answer and see that it is a reasonable consequence of Coulomb's law. First let us note that we do this work *very* slowly; no kinetic energy is given to the charged body, and the force we exert is equal to (*not* greater than) the electrostatic repulsion of the two charges.

The correct answer for the work we must do in moving q_1 from r' to r_0, keeping q_2 fixed, is

$$W = \frac{kq_1q_2}{r_0} - \frac{kq_1q_2}{r'} = \left(\frac{kq_1q_2}{r_0r'}\right)(r' - r_0) \tag{2.1}$$

Here, to save writing in dealing with point charges, we are letting k stand for the constant in Coulomb's law, as at the beginning of Chapter 1. Equation (2.1) is clearly correct if the displacement $(r' - r_0)$ is very small, so that $1/r_0r'$ is very nearly $1/r_0^2$ or $1/r'^2$. It can be justified for larger displacements by plotting the Coulomb law force against variable r, and finding the area under the curve between any two particular values of r (Figure 2.2), just as one finds the work done by (or against) a mechanical force which is not constant over the dis placement. Actually it is valid for any displacement, large or small, of along the line on which q_1 and q_2 lie.

If we begin with q_1 so far away from q_2 that the force on it is negligible, the outside work required to bring it up to distance r_0 from q_2 is simply

$$W = \frac{kq_1q_2}{r_0} \tag{2.2}$$

Once the work has been done we may say that the pair of charges themselves possess this energy. We have held q_2 fixed while bringing

up q_1, so that we could say, alternatively, that q_1 now possesses energy kq_1q_2/r_0 owing to its position with respect to q_2, energy which it did not have when it was far away.

We know from mechanics that energy a body possesses by virtue of its position is called *potential energy*. If we say that kq_1q_2/r_0 is the potential energy of qt at distance r_0 from q_2, we are calling the potential energy of qt zero for $r = \infty$. This is an arbitrary but convenient floor from which to measure electric potential energy of point charges.

The work represented by (2.1) depends only on the *radial* displacement. That is because the force be tween q_1 and q_2 is along the line joining them; the charge q_1 could be moved anywhere on a sphere of radius r_0 about q_2 as a center without costing any work whatever against electric forces (Figure 2.3). Moreover, in moving the charge q_1 from P' to P_0 in Figure 2.4, the same amount of work is done for all the paths shown, only the radial portions of the path require any "pushing" against the repulsion of the charges, and in doubling back to larger distances from q_2 one gains energy from the repulsion. The work put into carry ing charge q_1 from P' to P0 is i $kq_1q_2(1/r_0 - 1/r')$, regardless of the path. Forces for which the work done in going from one point to another is independent of the path are called *conservative* forces, and electrostatic forces are conservative.

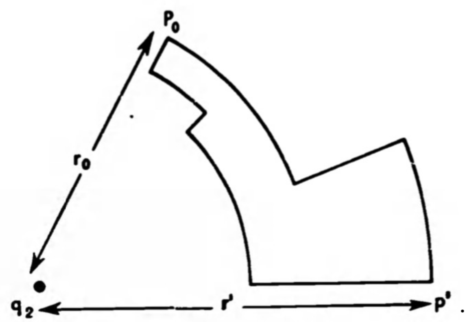


Figure 2.4: Work done by the moving charge rom P' to P_0.

Let us rewrite the external work required to move q_1 from r' to r_0 in terms of the electric field intensity associated with the charge q_2 as given by (1.3) of Chapter 1. From the definition of work as the scalar prouct of force and displacement, we write the work done against the force $q_1\vec{E}$, for a displacement Δs, so small that $\vec{E}$ is essentially constant over the distance interval, as $q_1\vec{E} \cdot \Delta\vec{s}$. It is exactly the sum of such increments of work that we have considered in obtaining (2.1).

$$\begin{aligned} W = \sum -q_1\vec{E} \cdot \Delta\vec{S} &= q_1\left(\frac{kq_2}{r_0} - \frac{kq_2}{r'}\right) \\ &= q_1(\Phi_0 - \Phi') \end{aligned} \tag{2.3}$$

where the sum is over all As on any path from P' to P_0. The last expression for this work makes explicit its dependence on q_1 and the difference of two quantities which depend on the end points of the path in relation to the source of field intensity, q_2. The quantity $(\Phi_0 - \Phi')$ is called the *difference of potential* between points P_0 and

P'. It is the *work per unit positive charge* required for the transfer of position from P' to P_0.

The difference of potential between any two points in an electrostatic field may be defined as the *external work per unit positive charge* required to move the charge from one point to the other, regardless of whether the field is that of a point charge. In going from point P_1 to point P_2,

$$\Phi_2 - \Phi_1 = -\sum \vec{E} \cdot \Delta\vec{S} \tag{2.4}$$

where the sum is to be taken over all line elements of any path beginning at P_1 and ending at P_2. The negative sign is due to the fact that in doing work on the positive charge you are acting against the field if the charge has more potential energy at the end of the path than at the beginning. If $(\Phi_2 - \Phi_1)$ is positive, we say that P_2, the final point, is at a higher potential than P_1, the initial point.

In order to use the word *potential*, instead of *potential difference*, we must establish some conventional floor, just as for potential energy. What floor is established, what position is said to be at zero potential, is a matter of convenience. For a point charge, or a configuration of charges localized in a small region of space, it is customary to say that the potential is zero at very large distances from the charge, i.e., at $r = \infty$. Thus the potential at point P which is a distance r from point charge Q is

$$\Phi = \frac{kQ}{r} \tag{2.5}$$

Since, as we have seen, the field in tensity outside any distribution of charge having spherical symmetry is indistinguishable from that of a point charge at the center of the sphere, (2.5) is also the potential at a point r distant from the center of a sphere of charge, so long as the point is outside the sphere.

The law of superposition holds for potential as well as for field intensity, and the superposition is much easier to accomplish for a scalar quantity than for a vector field. The potential at point P in Figure 2.5 is simply due to the presence of three fixed charges, $k(q_1/r_1 + q_2/r_2 + q_3/r_3)$ It is left to the problems to show that, with the signs and magnitudes of charges indicated, the potential at P is positive, while the potential at P' is zero. This means that work would have to be done *by* some outside force to bring a positive charge from great distance to P, whereas no net work is needed to bring one to P'. Other points in the field are at negative potential. What does *that* mean?

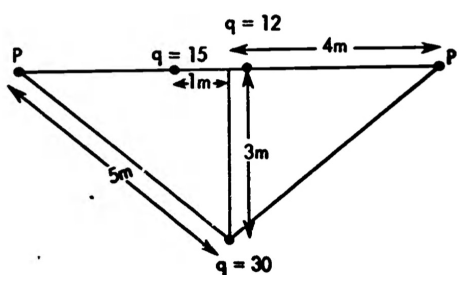


Figure 2.5: Law of superposition holds for potential.

The applicability of the law of superposition means that a general formula can be written for the potential whose source is a fixed distribution of charge:

$$\Phi = \sum \vec{E} \frac{k\,\Delta q}{r} \tag{2.6}$$

is the ordinary numerical sum over all elements Δq of the distribution, each with its own r. This equation can be compared with (1.2) of the previous chapter; it is much easier to evaluate the potential in many problems, simply because the potential is a scalar.

But of what use is it to know the potential? Presumably we can measure forces on charges with torsion balances or otherwise, and the mapping of electric field lines, as in Chapter 1, tells us a great deal about such forces, but measuring the total work in ringing a test charge from infinity to each point is merely a "thought experiment." Thought experiments are often very useful in understanding physics, but the potential has a more practical value as well. Even in very complicated applications, such as the electrostatic electron microscope, it is often the potential that is mapped, rather than the field itself.

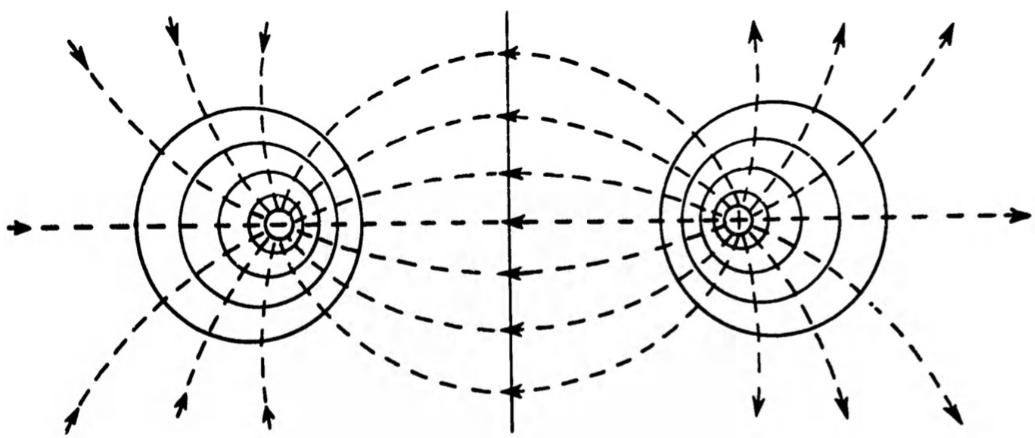


Figure 2.6: Equipotential lines between two equal and opposite charges.

The potential is mapped by drawing *equipotential lines*, or the traces of *equipotential surfaces*. The device is familiar from contour maps showing the altitude (gravitational potential) of a geographical region, but that is a very simple example, since the gravitational potential varies only with the vertical coordinate of space, while electric potentials vary with three coordinates. The equipotential

surfaces corresponding to the two equal and opposite point charges as cut by a plane containing the charges are shown in Figure 2.6. In this instance the equipotentials are surfaces of revolution about an axis on which the charges lie. The field intensity lines are also sketched; they must meet the equipotentials at right angles, since by definition no work is done against electrical forces in moving from one part of an equipotential surface to another.

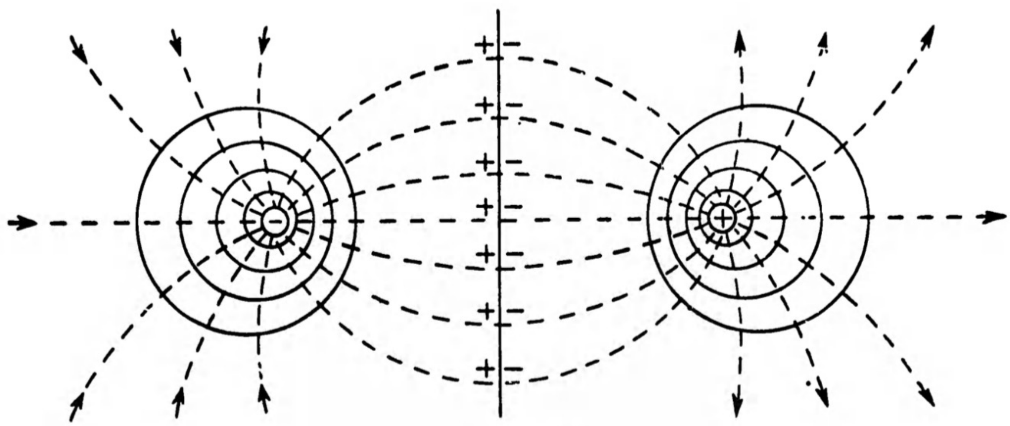


Figure 2.7: Equipotential lines between two equal and opposite charges do not change with a plane conductor between them.

The requirement that field lines be orthogonal to equipotentials enables us to map the field, knowing the equipotentials. For some charge configurations the potential may be computed (or determined in some other way), equipotentials mapped, and the field lines are ascertained. When conductors are involved, the potential plays a particularly simple and important role.

In Chapter 1 we defined a conductor as an object in which charge is free to move, and noted that almost by definition there is no electric field within the substance of a conductor in electrostatics. It follows equally that there is no difference of potential between different points in or on a conductor if the charge is static. The surface of a conductor *is* an equipotential surface in electrostatics – in fact the entire volume of a conductor is a single potential. If a thin sheet of copper, for example, were substituted for the equipotential plane of Figure 2.6, as indicated in Figure 2.7, the field on both sides of the plane would be completely unchanged. The field lines do net penetrate the conductor, to be sure; free charges appear on the conductor in just such a configuration that the plane remains an equipotential. But the conductor is also an electrostatic shield: One of the two point charges could be removed, and the field on the

other side would remain just as before. The charge on the conductor is said to be *induced* by the presence of the point charge Q.

In general an uncharged conductor placed in an electrostatic field distorts the field so its surface *becomes* an equipotential (Figure 2.8). This process involves separation of the surface charge in such a way that the field in tensity normal to the surface has a magnitude $E = \sigma/\epsilon_0$, in accord with (1.11). Such charges are said to be *induced* on the surface of the conductor. The conductor may have a net charge; if so, the charge is redistributed by the presence of another charge nearby, so that the surface is at a single potential. The conductor may be "grounded," connected to the earth, which is a large and reasonably good conductor itself, a reservoir of charge which may be considered infinite Figure 2.9. The potential of earth, or any conductor in electrical contact with it, is said to be zero; is this consistent with the zero potential assumed for an isolated point charge?

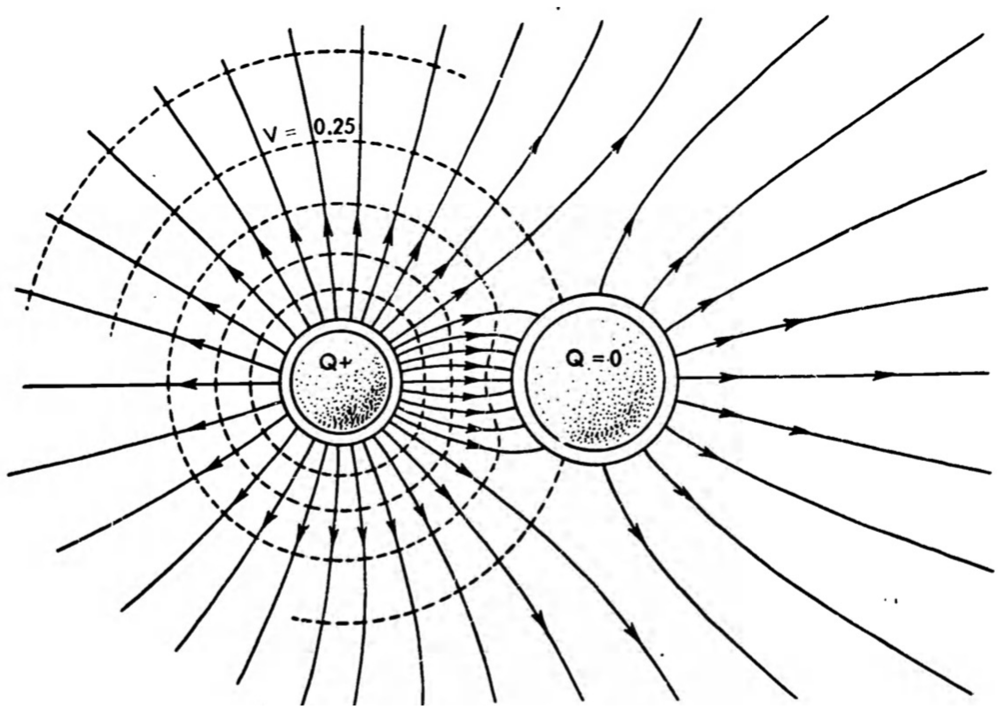


Figure 2.8: Uncharged conductor in electrostatic field has an equipotential surface.

In configurations of conductors, difference of potential is often more important than any attempt to ascribe potential itself. A particularly simple and instructive set of conductors consists of a pair of parallel plates of equal area, separated by a distance d small com-

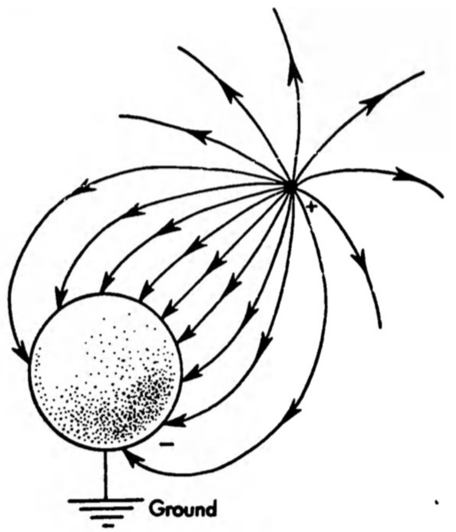


Figure 2.9: Potential lines between a grounded conductor and an isolated charge.

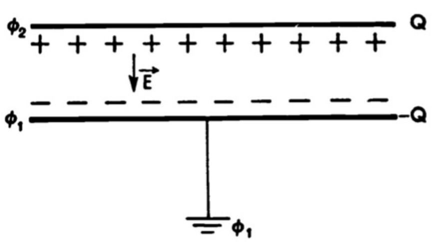


Figure 2.10: Potential between two plates of equal area separated by a small distance.

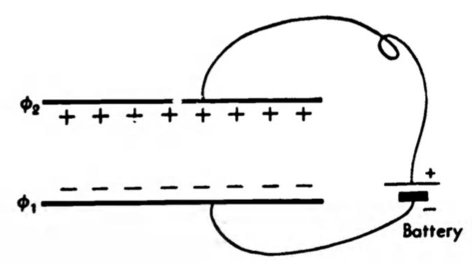


Figure 2.11: Potential between two oppositely charged plates by a battery of equal area separated by a small distance.

pared with the dimensions, of the plates (Figure 2.10). Throughout most of the region between the plates the fact that there are edges can be ignored. If the upper plate is given a charge +Q, the field is $\vec{E} = \sigma\,\vec{n}/\epsilon_0$, where $\sigma = Q/A$, and the equipotential surfaces are planes, as indicated, from the symmetry of the configuration. The equipotential surface which is the lower plate has an induced charge –Q. The field between the plates is uniform and the difference of potential between the plates is $\Phi_2 - \Phi = \sum \vec{E} \cdot \Delta\vec{S}$. In scalar magnitude,

$$\Phi_2 - \Phi_1 = E \cdot d = \frac{Q\,d}{\epsilon_0 A} \tag{2.7}$$

The MKS unit of potential difference is the volt, defined as one joule per coulomb: A potential difference of 1 volt means 1 joule of work against electrical forces to move unit charge from one point to the other. From equation (2.7) it is easy to see why field in tensity is often expressed in volts/ meter instead of newtons/coulomb.

Instead of stroking the top plate with a glass rod which has been rubbed with silk and letting a charge be induced on the lower plate, it is more convenient to connect the two plates to some source of difference of potential by means of conducting wires. The most familiar source of potential difference is the chemical cell, or battery. A battery may be defined for our purpose as a device capable of maintaining a constant potential difference between two electrical "terminals," and of supplying equal and opposite amounts of charge when necessary to accomplish this result. The chemical battery was invented by Alessandro Volta of the University of Padua, Italy, in 1800, and greatly facilitated the growth of electrical science.

A source of potential difference is actually a source of energy, a de vice for transforming energy of some other kind into electric energy. The amount of energy per unit charge it is capable of delivering is called its *emf* – the letters stand for electromotive force, but the name is an anachronism, since *emf* is not a force. The potential difference between the terminals of a battery, or between the plates of Figure 2.11, is *equal* to the *emf* of the battery, but an *emf* cannot be produced electrostatically.

To find the energy required to charge the plates of Figure 2.10 or Figure 2.11, we note from equation (2.7) that the amount of charge on either plate is directly proportional to the potential difference between the plates. This is usually written

$$Q = C\,(\Phi_2 - \Phi_1) = C\,V \tag{2.8}$$

where C is a constant independent of the charge or potential, and $V = (\Phi_2 - \Phi_1)$ to save writing. Let us begin with uncharged plates, and transfer positive charge Δq from the lower plate to the other, leaving $-\Delta q$ behind. The increment of work ΔW required to trans port any later Δq is proportional to the potential difference, which builds up as the charge on the plates in creases, and is equal to q/C for any net plate charge q.

If the potential difference is plotted against q as in Figure 2.12, and the product $(q/C)\Delta q$ is summed for all Δq from $q = 0$ to $q = Q$ we obtain the total work done in charging the plates; it is equal to the area under the straight line, which may be written in several ways,

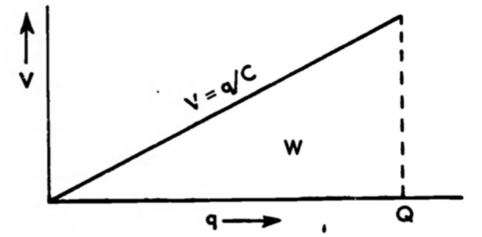


Figure 2.12: Potential difference as a function of charge q between two oppositely charged plates.

$$\begin{aligned} W &= \frac{1}{2}QV = \frac{1}{2}\frac{Q^2}{V} \\ &= \frac{1}{2}\frac{Q^2 d}{A\epsilon_0} = \frac{1}{2}QEd \\ &= \frac{1}{2}\epsilon_0 E^2 A d \end{aligned} \tag{2.9}$$

by virtue of equations (2.7) and (2.8). This energy is stored in the configuration of charged conductors, but an interest ing alternative interpretation is suggested by the last form given in (2.9). The product of the area A and the plate separation d is just the volume of space between the plates where the field intensity is $\vec{E}$. Outside this space the field vanishes, if we neglect edge effects. We can, if we like, at tribute the energy to the field itself, an energy equivalent to $E^2/2\epsilon_0$ joules/ cubic meter. The question of whether the energy *really* resides in the charged conductors or in the space between them has a mechanical analogy: In a loaded air rifle does the bullet have potential energy, or is the potential energy in the compressed air?

Since the plates of Figure 2.11 are oppositely charged, they attract each other, and mechanical forces are required to hold them in place. From equation (2.9) we may determine the amount of force required. Let us consider the work necessary to increase the separation by a small distance Δd, without changing the charge on either plate.

$$W = F\,\Delta d = \frac{1}{2}Q\,\Delta V = \frac{1}{2}QE\,\Delta d$$

so that the external mechanical force needed may be identified as

the coefficient of the small displacement Δd,

$$\begin{aligned} F &= \frac{1}{2} Q E = \frac{1}{2} \epsilon_0 E^2 A \\ &= \frac{1}{2} \frac{Q^2}{A \epsilon_0} \end{aligned} \tag{2.10}$$

which, like the amount of stored energy, may be expressed in a variety of ways.

A configuration of two conductors near each other but not in electrical contact is called a capacitor or a condenser. The proximity of charges of the opposite sign makes it possible to accumulate relatively large charges. The first man-made capacitors were called Leyden jars, and consisted of glass jars with a conducting substance inside and a conductor outside. A jar of water was held in the hand, according to the first record of the device, and a conducting wire connected the water to an "electrical machine" capable of supplying charge. The stored energy became so great that when the other hand was brought near the wire there was a discharge through the air in the form of a spark. To Franklin, who repeated such experiments, the spark was reminiscent of lightning. It was this idea that he tested in his famous experiments with kits in thunderstorms, and so established the connection between lightning and electricity.

Exactly how the clouds become charged is a very difficult and complicated problem, but they do, and they are enormous conductors, near a still larger one, the earth. The discharging spark, which releases the gradually accumulated electric energy, constitutes lightning. Differences of potential be tween cloud and ground may be as high as one billion (10^9) volts, and a discharge of 20 coulombs is not unusual, so that the energy dissipated in a single stroke of lightning may be 10^{10} joules, equivalent to nearly 3000 kilo watt hours. The discharge itself is of course not an electrostatic phenomenon, but its existence is evidence that electric forces are indeed strong.

Chapter 3

Electrical Properties Of Material Media

In our consideration of the effects of electric charge we have begun as if there were no intervening matter be tween charges. The electric field in tensity $\vec{E}$ (force per unit charge) is defined and traced to its sources as if the charges existed in vacuum. On the atomic level, this view can be maintained and justified, since we hold that matter consists of "atoms and void." But in practice we experience gross matter, many of whose properties cannot be traced in any simple way to the behavior of the atoms and subatomic constituents. In fact, many properties of matter have not to date been satisfactorily traced to the behavior of its constituent particles, even if all known complications are taken into account. These problems are the subject of intensive on-going research.

We shall not here restrict our selves entirely to the gross aspects of phenomena, for a number of them can be understood quite simply, if only qualitatively, in terms of atoms and molecules. Nevertheless we shall begin with the consideration of matter in bulk.

Electrical effects were first observed and studied with what were once called "electrics" and now called insulators. For quantitative observations on static charges, however, it is necessary to use conductors, typically metals, on the surface of which net charge is distributed, and which will maintain their charge if well insulated. Experiments such as Coulomb's car be performed in dry air or in an evacuated vessel with no very appreciable change in the results, and except as convenient supports for charged metal objects, the role of insulators in the science of electricity was minor. (Insulators are, of course, also important for support of current-carrying con-

ductors, but again for their *lack* of electrical properties, that is, as *nonconductors* of electric charge.)

But insulators do have electrical effects, even when they possess no net charge. A systematic investigation of these effects was made by Faraday. His experiments were varied, and often complicated, but the essential results may be inferred from consideration of very simple apparatus. Let us consider a pair of fixed parallel metal plates, large in comparison with the separation between them, which can be charged by connecting across a battery, as indicated in Figure 3.1.

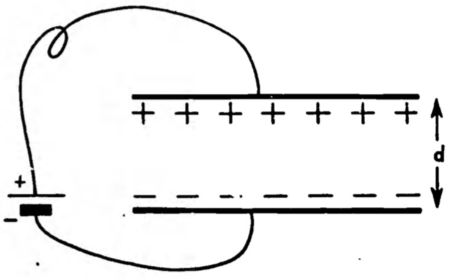


Figure 3.1: Field intensity between two plates charged by a battery.

The battery imposes a known difference of potential between the plates, which can be measured with a good voltmeter. Now disconnect the battery by opening the switch, leaving the plates charged and well insulated. The space between the plates is vacant (dry air is very nearly equivalent) and the presence of the charge pro duces a field intensity $E = Q/A\epsilon_0 = \sigma/\epsilon_0$ in that region. Here Q is the total charge per plate of area A, and $\sigma = Q/A$ is the surface density of charge. We are assuming the plates so large that edge effects can be neglected. The difference of potential between the plates is just *Ed*, the work per unit charge in going from one plate to the other. This last result follows from the definition of potential difference, as force per unit charge times distance, both along the same direction.

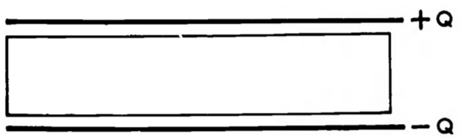


Figure 3.2: Field intensity between two plates charged by introducing a glass slab.

Now insert a slab of uncharged insulating material, such as glass, so as to very nearly fill the space be tween plates but, as a precaution, without touching them (see Figure 3.2). Note that the plates are charged just as before. If the difference of potential is now measured again, it is found to have changed. In fact, for glass nearly filling the space the voltage is reduced from its earlier value by more than a factor of two. For all isotropic insulating materials, the effect is to reduce the voltage, and Faraday found that the reduction factor is a constant for any particular material – it does not depend on the particular geometry of the system of conductors.

What must we conclude? The charge has remained unchanged, but if we are to retain the relation between field intensity and potential difference, which follows from the definitions of force and work, we can only conclude that the field intensity E in *Ed* has changed within the material. Faraday called such materials *dielectrics*, each with its own dielectric constant; if the voltage is reduced by a factor of 2, the dielectric constant is 2, for example. We should note that most materials, including glass, are found to be entirely unchanged when removed from the apparatus.

This effect of the dielectric can be described in terms of its

polarizability. Figure 3.3 shows only lines of E, and indicates that there are sources (or "sinks") of $\vec{E}$ on the surface of the slab, even though its *net* charge is zero. This would account for a reduction in the strength of E inside the dielectric, as a factor in $V = Ed$, the potential difference. What has happened to the dielectric itself? Let us consider the slab alone; the effect of the charged plates remains, but the plates themselves are not shown in Figure 3.4.

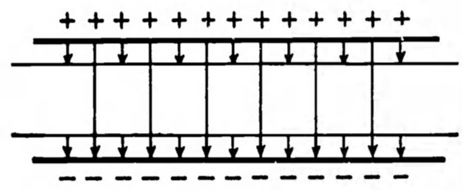


Figure 3.3: Field intensity lines between two plates indicating that there are sources (or "sinks").

The appearance of equal and opposite charge on the two flat surfaces would arise if the normal dielectric consisted of equal and opposite charge densities occupying the same volume and thus cancelling each other, but now one kind of charge is slightly displaced relative to the other. The result is charge neutrality except at the faces perpendicular to the relative displacement, so that each such face has a surface charge. Now a familiar configuration of two equal and opposite point charges is called an electric dipole; its strength, or dipole moment, is the product of the magnitude of one charge! and the distance be between them. The dipole moment is actually a vector quantity, whose direction is taken from the negative toward the positive charge as indicated in Figure 3.5.

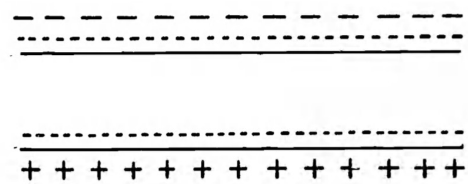


Figure 3.4: The change in dielectric due to electric field.

The slab of dielectric in the previous figure is clearly a dipole, extending throughout a volume instead of being merely a line. If we let the charge density on the surface of the plate be σ_p, so that the total charge is $Q_p = \sigma_p A$, the dipole moment of the slab is $\sigma_p Ad$, where d is the thickness. It is directed from negative $-\sigma_p$ to positive $+\sigma_p$. But *Ad* is the volume of the slab, and we may define a dipole moment per unit volume $\vec{P}$, whose magnitude at the surface perpendicular to $\vec{P}$ is σ_p.

We note that the polarization extends throughout the volume, but since it is uniform the only accumulation of charge is at tho surface. Even there the charge is not *accessible*; it is firmly attached to the body of the material, and results only from the slight displacement of positive from negative charge in neutral matter.

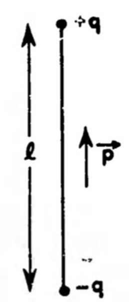


Figure 3.5: The dipole moment between two charges.

The dipole moment per unit volume of a dielectric can be equally well attributed to the atoms or molecules of which it is composed. Suppose that every atom is permanently neutral, but that the positive charge is displaced slightly from the negative charge. Thus every atom is a small dipole, whose dipole moment is $\vec{p}$. If there are N such dipoles per unit volume, all in the same direction $\vec{P} = N\vec{p}$ is an equivalent description per unit volume. We do not explain anything by putting the volume polarization in this form, but we do note that a uniform continuous distribution of charge is

not necessary for its definition.

In the case we are considering, with a dielectric such as glass, the polarization exists only in the presence of an electric field $\vec{E}$, and is in fact proportional to $\vec{E}$ over wide range, but some materials have intrinsic polarizations, and the proportionality of $\vec{P}$ and $\vec{E}$ is not a fundamental fact of electricity. Our description applies both to intrinsic polarization and that produced by the presence of an electric field. Let us note the general relation between the volume polarization of matter and the polarization charge which appears on its surface.

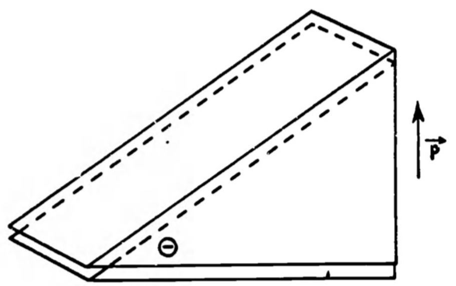


Figure 3.6: The dipole moment of a wedge of a dielectric.

The wedge of dielectric in Figure 3.6 has a cross section which is a right triangle, and is polarized in the direction of one leg. This polarization can be thought of as produced by displacing vertically the whole volume of positive charge from the same volume of negative charge. Charge thus appears on only two surfaces, the bask and the slant face, equal and opposite t in total amounts. At the base, to which $\vec{P}$ is normal, we can find that σ_p is equal to the magnitude of $\vec{P}$, by the same arguments developed in connection with the parallel slab. But the area of the slant face is greater than that of the base: $A_{\text{slant}} \cos\theta = A_{\text{base}}$, where θ is the angle between the base and hypotenuse of the cross section.

It follows that the surface density of polarization charge on the slant face is smaller than that on the base, i.e., is equal to $P\cos\theta$, since the total charge on the two faces is the same in magnitude. The quantity $P\cos\theta$ is simply the normal component of $\vec{P}$; in general

$$\sigma_p = P_\perp \tag{3.1}$$

at the plane face of any polarized dielectric.

Since the surface charge on a dielectric is truly inaccessible to direct measurement, as would also be variations in $\vec{P}$ within the dielectric, it is advantageous to write the description of electrical phenomena directly in terms of $\vec{P}$ and accessible (separable) charge. Let us write

$$\vec{D} = \epsilon_0\vec{E} + \vec{P} \tag{3.2}$$

where $\vec{P}$ and $\vec{E}$ have now been defined. The new field quantity, $\vec{D}$, has as sources in terms of Gauss's law only accessible charge:

$$\Phi = \sum_{\text{closed}} D_\perp\,\Delta S = Q_{\text{accessible}} \tag{3.3}$$

the *net* free charge contained within the volume. Lines of $\vec{D}$ thus begin and end only on separated charges; unlike those of $\vec{E}$, they are continuous through the surface of a polarized dielectric.

For historical reasons which go back to the idea of an ether pervading all space and having mechanical properties of its own, $\vec{D}$, is called the displacement field. It is defined here in terms of two quantities which have more immediate physical significance, the electric field intensity, $\vec{E}$, and the electric dipole moment per unit volume, $\vec{P}$, of dielectric material In empty space $\vec{D} = \epsilon_0\vec{E}$, and within a dielectric we could make a complete description of electrical effects in terms of $\vec{E}$ and $\vec{P}$ alone. But we shall see at once that the combination of $\vec{E}$ and $\vec{P}$ given by equation (3.2) is very convenient for expressing electric energy.

The electric energy stored in a pair of charged parallel conducting plates at distance d from each other., given by (2.9) when there is empty space between the plates, may also be computed for plates separated by a dielectric. The basic equation, $W = QV/2$, where Q is the magnitude of the charge on each plate and V is the difference of potential between the plates, is still valid, but $\vec{E}$ is no longer so simply related to the charge on the plates. Application of Gauss's law to the surface of a conductor yields

$$\vec{D} = \sigma_{\text{accessible}}\vec{\hat{n}} = \frac{Q}{A}\vec{\hat{n}} \tag{3.4}$$

where $\vec{\hat{n}}$ is a unit vector normal to the surface, and $\sigma = Q/A$ for our uniformly charged plates. Equation (3.4) holds whether there is empty space or some dielectric substance outside the surface of the conductors, but $E \neq \sigma/\epsilon_0$ within a dielectric. Since $V = Ed$, however,

$$W = \frac{1}{2}QV = \frac{1}{2}DEAd \tag{3.5}$$

for the energy stored in the parallel plate configuration of charges. We note that this expression is equally valid whether the space between the plates is empty or filled with a homogeneous dielectric substance.

The problem of forces between charged bodies embedded in a dielectric must be approached with considerable care. If we knew the volume polarization everywhere, as well as the positions and magnitudes of all free charges, we could in principle apply Coulomb's law directly to find the forces due to all charges, including those which appear on or in the dielectric medium as a result of its

polarization. Polarization is usually *induced*, depending for its very existence on the presence of fields produced by accessible charges. For isotropic homogeneous materials we have already noted that the dipole moment per unit volume, $\vec{P}$, is directly proportional to the field intensity $\vec{E}$. For such materials $\vec{D}$ as given by (3.2) is then also directly proportional to $\vec{E}$, and the factor κ in the relation

$$\vec{D} = \kappa \epsilon_0 \vec{E} \tag{3.6}$$

is called the *dielectric constant* of the material. The dielectric constant is a pure number, found empirically to be greater than 1 for all substances. If this relation is applicable, a simple expression for the forces between charged conductors may be derived.

Since the mechanical forces which account for the rigidity of solids may be very complicated, let us consider the forces between two charged parallel plates, as in Figure 3.1, when immersed in a fluid dielectric. A detailed calculation of the forces would include changes in fluid pressure produced by polarization of the fluid, but we may find the total force of attraction between the plates from the expression for electric energy, just as we arrived at equation (2.10) in Chapter 2. To increase the separation of the plates by a small distance Δd without changing the charges on the plates we must do an amount of work

$$W = F\Delta d = \frac{1}{2} DA\Delta d = \frac{Q^2}{kA\epsilon_0}\Delta d$$

where the last expression follows from equations (3.4) and (3.6). Therefore

$$F = \frac{Q^2}{\kappa A \epsilon_0} \tag{3.7}$$

for the force of attraction between two charged plates of area A immersed in a fluid dielectric whose dielectric constant is κ. The result differs from the force between the plates in empty space by the factor κ in the denominator. Since $\kappa > 1$, the force between the plates is reduced from its vacuum value.

Net forces between charged bodies immersed in a fluid dielectric are always found to be reduced by a factor κ in comparison with the force in empty space. This result is consistent with the fact that $\vec{E}$, the electric field in tensity within the dielectric, is also reduced in a homogeneous dielectric from the value it would have in empty space, but the change in the force comes as a result of

induced polarization charges in or at the boundaries of the dielectric. Coulomb's law itself is not changed by the presence of the dielectric. The direct electrical interaction between two charges remains the same, but other sources of electric field have been created by the polarization of the dielectric.

Electrostatic forces play an extremely important role in nature. They account for the binding of electrons in atoms, and the binding of atoms into molecules, although the details of such "accounting" are very complicated indeed. Other important aspects of electricity that appear when charges are in motion are the subject of Monographs II and III of this series.

Problem

3.1 The lightweight objects attracted by electrified amber or glass are bits of dielectric polarized by the presence of the electric field.

   (a) Describe qualitatively the distribution of charge in such a small object while it is being held by a negatively charged piece of amber.

   (b) Is there a net force on a polarized dielectric in a *uniform* electric field?

   (c) Some dielectric substances have an intrinsic electric dipole moment per unit volume; a body possessing this property is called an *electret*. Would a small electret be attracted to a piece of electrified amber or glass? How could you distinguish between electrets and bits of dielectric which have no permanent polarization?

Chapter 4

Electrostatics Reformulated

Coulomb's law is all there is to electrostatics, but many aspects and con sequences of the law appear clearly only when the subject is formulated somewhat more mathematically. For students wishing to pursue the subject further we recommend Volume II, Chapters 1, 2, 3, and 9 of the Berkeley Physics Course, written by E.M. Purcell. Professor Purcell's approach is "microscopic," based on a qualitative description of the electrical properties of atoms and molecules whose space average yields the field quantities $\vec{E}$ and $\vec{B}$. The theorems which emerge so elegantly from use of the calculus apply equally to continuous distributions of charge or to space averages of atomic charges: They depend essentially on the inverse square law and the superposition principle.

Part II

Magnetostatics

Chapter 5

Magnets And Magnetism

Magnetic iron ore, known to us as lode-stone or magnetite, is found in many parts of the world, and its property of attracting iron was noted by more than one civilization early in its Iron Age. The property remained merely a curiosity, even in the intellectual climate of the Golden Age of Greece, and the tendency of a magnet to orient itself along the earth's meridian escaped notice. These directive effects were probably first discovered in China, but there is no conclusive evidence that they were put to practical use. The origin of the mariner's compass is shrouded in mystery which may never be dispelled, but by the end of the 12th century the compass was well known in the Western world as a helpful device for sailors when the stars wore obscured. Myths, legends, and superstitions about magnets multiplied from ancient times through the Middle Ages, and even later. Magnets were employed in medicine, especially for the healing of wounds, and once the directional properties were recognized, in the occult sciences such as astrology. Yet magnetism began to be a genuine science during the Middle Ages; The first account of the magnet that we would call scientific was surprisingly early, a letter dated August 12, 1269. The epistle of Peter Perigrinus (Peter the Pilgrim), born Pierre de Marlcourt in Picardy, sets forth a number of fundamental properties of magnetism.

It was Perigrinus who discovered poles and distinguished precisely two kinds. His method is of interest; Select a good piece of magnetite, shape it into a sphere and polish it. Now place on it a needle or sliver of iron, and mark on the surface of the sphere the direction taken by the needle, Repeat the procedure at many different positions on the sphere. At the end it is found that the lines "will run together in two points, Just as all the meridian circles of the

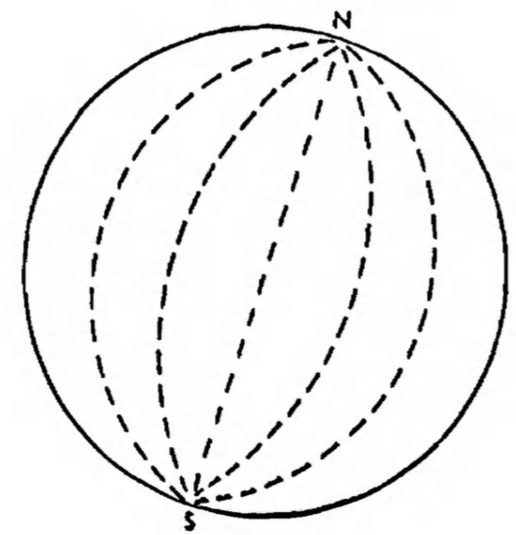


Figure 5.1: The lines of meridian on the globe.

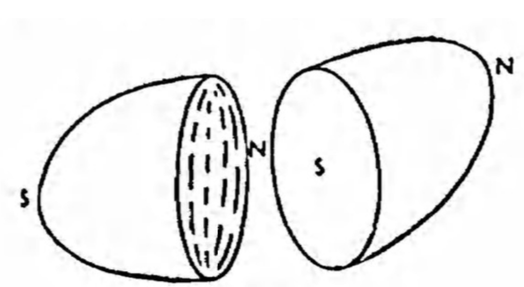


Figure 5.2: A magnet broken into two exhibits further poles.

world run together in two opposite poles of the world" (Figure 5.1). Only one of these poles points north if the magnet is free to turn. Thus two opposite magnetic polos woe introduced, and Perigrinus noted that unlike poles attract, lie went further to show that if a magnet is cut (see Figure 5.2), two poles persist in every separated part, and that if two fragments are put together as before the new poles vanish. He did not notice repulsion. Perigrinus named the poles north and south, with the north pole that which points to the north. At that time magnetism was attributed to the celestlal sphere, not to the earth itself, and Perlgrlnus did not think of the earth as a magnet.

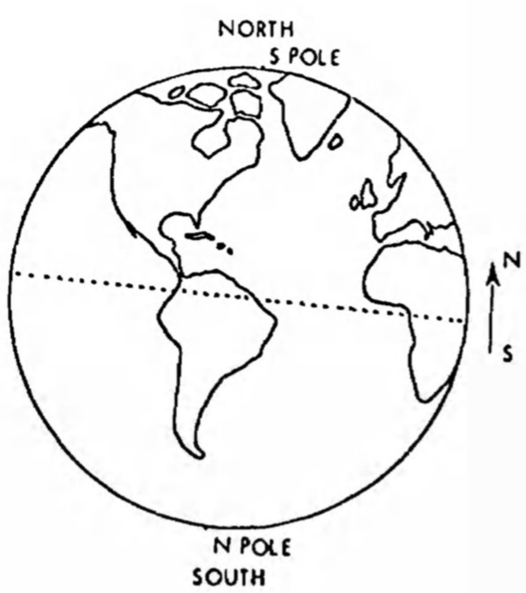


Figure 5.3: The Earth as a magnet.

This last step was taken by William Gilbert, physician to Queen Elizabeth, who repeated and extended the experiments of Perigrinus and others. Once the earth is taken as a magnet, the action of a compass is simply an example of the behavior of all magnets: The north magnetic pole of the earth is a "south" pole, which attracts the "north" poles of all compass needles (see Figure 5.3). Gilbert's great accomplishment was to detract existing facts anti laws of magnet ism from a wealth of speculation and superstition, and to discover new properties a relations. His theories, including the supposed relation of magnetism to gravitation, need not concern us here, although we should note that they influenced Kepler, and that Newton found them suggestive in the development of his own ideas. Gilbert's book *De Magnete*, published in 1600, is still a classic presentation of many qualitative aspects of magnetism.

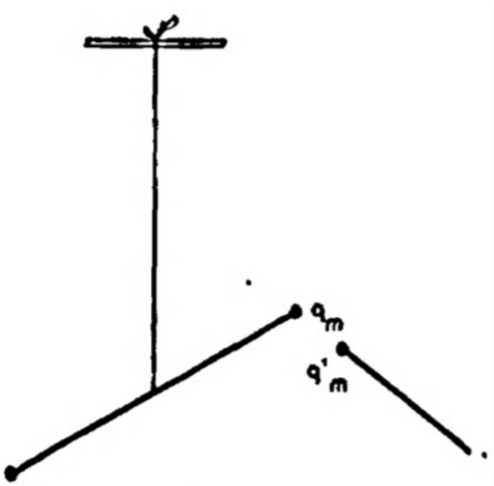


Figure 5.4: The magnetic effect is concentrated at the ends of a magnet.

Magnets exert forces on each other and on iron without being in contact. It was known to Gilbert and those who followed him that the effect of a magnet decreases as the distance from the magnet increases, but the quantitative relationship was first discovered by the Reverend John Michel1 in 1750. The same relation was found by Coulomb in 1785. The torsion balance, which facilitated these experiments, was invented independently by Michell and by Coulomb, und Michell's balance was later used by Henry Cavendish for his famous measurement of the constant in Newton's law of gravitation.

In long magnetized needles, or stiff wires of hardened iron, the magnetic effect is well concentrated at the ends, or poles (Figure 5.4). Michell established experimentally that the two poles are opposite and of equal strength for any one magnet, and that repulsion and attraction between the poles of two magnets are of equal magnitude if distances of separation are kept the same. He also found that the force exerted by the pole of a long magnetized wire is the same

in all directions. He then determined the force between poles for various distances, and found its strength was inversely proportional to the square of the distance between poles.

In order to write the law of Michell and Coulomb in mathematical form, we must assume some quantitative measure of pole strength, and it is convenient to call north-seeking poles positive (+) and south-seeking poles negative (–), by analogy to the designation of electric charge as positive and negative. If we designate pole strength by q_m, the force between two poles of strengths q_m and $q_{m'}$ has a magnitude given by

$$F = \frac{k\, q_m q_{m'}}{r^2}$$

with k an arbitrary constant determined by the units employed. In this formula a negative force signifies attraction, but force is actually a vector quantity, and the force exerted on q_m by $q_{m'}$ at distance r may be written

$$\vec{F} = \frac{k\, q_m q_{m'} \vec{r}}{r^2}$$

where r is a unit vector in the direction from $q_{m'}$ toward q_m. The force is directed as in Figure 5.5 if q_m and $q_{m'}$ are of like sign.

This law has exactly the same form as Coulomb's law for the interaction of electric charges, and a formal analogy between the interaction of magnets and electrostatic interactions can be carried further. The *magnetic field intensity* may be defined at any point in the vicinity of a magnet as the force per unit positive pole at that point. The conventional designation for magnetic field intensity is $\vec{H}$, so that

$$\vec{F} = q_m \vec{H}$$

where $\vec{H}$ is the field intensity at the position of q_m. The field intensity produced by a single pole $q_{m'}$ at distance r from the pole is

$$\vec{H} = \left(\frac{k\, q_{m'}}{r^2} \right) \vec{r}$$

where the unit vector $\vec{r}$ is directed *from* $q_{m'}$ *to* the point where $\vec{H}$ is to be determined. If $q_{m'}$ is negative (an S pole) the field intensity $\vec{H}$ is then directed toward $q_{m'}$. The field intensity produced by two poles is the vector sum of the contributions of the two poles taken separately – the principle of superposition applies. The fact that for real magnets the poles are not points, that pole strength is actually

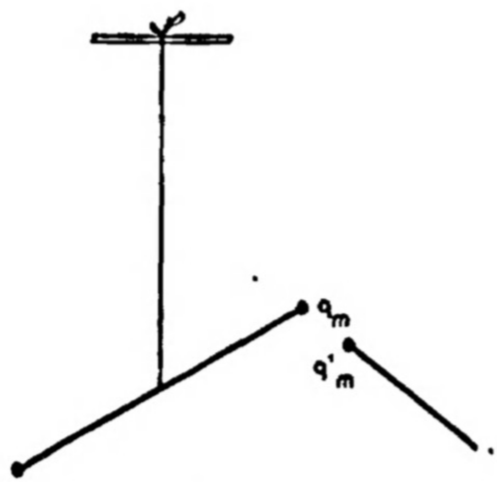
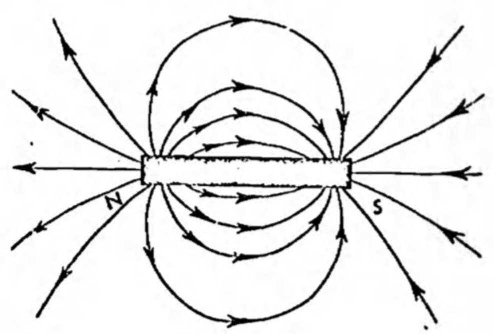


Figure 5.5: The force of attraction between magnets.

distributed over the surface or volume of a magnet, causes no difficulty: All effects are additive and can be summed over elementary surfaces or volumes at various distances from the point at which the field is to be computed, just as in electrostatics. In other words, the principle of superposition applies to magnetic forces, just as it does to electric forces.

We shall find that we have little occasion to work with poles as such in the further development of magnetism, although the magnetic field intensity remains an important concept, We should note that the MKS unit of pole strength is called the *weber*, so that $\vec{H}$ is in newtons per weber. (To anticipate Chapter 2: The weber is defined in terms of the ampere.) The constant k required to give the force in newtons if pole strength is measured in webers is $(10^7/16\pi^2)\ \mathrm{N\,m^2/Wb^2}$, or roughly 6×10^4. Clearly the MKS system is not designed to be convenient in working with poles, but the weber remains useful in other connections. It was named for Wilhelm Eduard Weber (1804–1891), who collaborated with the great mathematician Gauss in putting the whole question of electromagnetic units on a rational basis.



strength of magnetic field H.

Figure 5.7: The lines of field around a bar magnet.

In principle the magnetic field intensity in the vicinity of magnets could be computed at every point from a knowledge of the strength and position of all poles, but the numerical results would be hard to visualize. On the other hand, the direction and some notion of the strength of $\vec{H}$ can be demonstrated very easily with iron filings (Figure 5.6) or a large number of very small compasses. By definition a compass sets itself in the direction of the field intensity. A *field line* is a line so drawn that its tangent is in the direction of the field at each point, and to a good approximation a field line can be traced along its length by a small compass. The number of lines drawn does not matter, but we see from Figure 5.7 that the lines tend to converge in regions of high field intensity, and to become more widely spaced where the field intensity is weak. We are restricted to the region outside the magnet, and even then show only a two-dimensional cross section of space, but it is clear that the lines traced out in this way are smooth and continuous.

Iron filings and small compass needles are not entirely equivalent, although both are oriented along magnetic field lines. Soft iron, of which the filings are made, shows little or no *residual* magnetism; a small sliver of iron has no poles of its own, and orients itself equally readily if turned through 180° in the field of a magnet. The magnetism a piece of iron exhibits owing to the presence of magnetite or other permanent magnets is said to be *induced*.

The tiny compass needles with which lines of magnetic field intensity can be traced out are each complete with two opposite poles of equal magnitude. That the tiny magnet tends to align itself along the lines of $\vec{H}$ is to be expected, since the force on the positive pole is equal and opposite to the force on the nearly coincident negative pole. But poles are well localized only on long needles such as those used by Michel1 and Coulomb. (Coulomb's needles were 25 inches long.) For short needles it is practically impossible to determine a point position equivalent to the actual distribution of pole strength. A more convenient property by which the strength of a magnet can be measured is its *magnetic moment*, or, more precisely, its magnetic *dipole* moment. The dipole moment can be determined even for magnets which are entirely inaccessible for direct examination.

The dipole moment of a long magnet with poles at the ends is defined as a vector directed from the negative toward the positive (north-seeking) pole, whose magnitude is the pole strength q_m times the distance between the poles. Let us call the magnetic moment of such a magnet $\vec{m}$. If the magnet is placed in a uniform field, as indicated in Figure 5.8, the forces on it are equal and opposite, but it will experience a torque of magnitude $mH\sin\theta$, which tends to bring it into alignment with the lines of $\vec{H}$, (Here θ is the angle between the dipole moment $\vec{m}$ and the direction of $\vec{H}$.) For macroscopic magnets this torque can be measured oven if the pole strength is so diffuse that its exact position has little or no meaning. For microscopic linear magnets the dipole moment is the only accessible measure of magnetic strength.

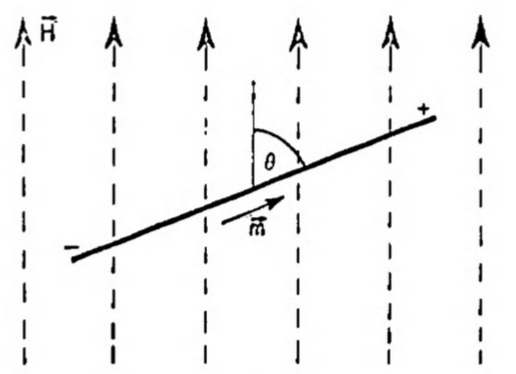


Figure 5.8: Magnetic moment in an uniform field.

Even the torque may be difficult to measure for a magnet so small as to be mechanically inaccessible with a device such as a torsion balance. The dipole moment can still be determined by finding the energy involved in lining up such magnets in the direction of the field lines, or by finding the energy required to reverse their positions. We note that if the magnet is initially in the direction of if, the work required to turn it through 90°, so that it becomes perpendicular to the field, is mH. (Proof of this statement is left to the problems.) It is conventional to say that the magnetic dipole has zero energy when it is perpendicular to the lines of $\vec{H}$. The energy is equal to $-mH\cos\theta = -\vec{m}\cdot\vec{H}$ for any other orientation. This is potential energy, since it is determined by the position of the magnet relative to the field, and the negative sign makes the most stable position (alignment *with* the field lines) that of the lowest

potential energy.

The description of magnetic interactions in terms of an inverse square law of force between poles seems to go very smoothly, but it has a distinct weakness: It does not include any statement of the experimental fact that poles cannot be isolated from each other. This could, of course, be added in words. Coulomb went further than Michell: After establishing the inverse square law and the direct dependence on magnetic pole strength, he postulated that the "molecules of magnetic fluid" are themselves elementary magnets, complete with two poles of equal strength. Chains of such "molecules" would then cancel each other except at the ends, which appear as magnetic poles. This would explain the fact that two poles appear when a magnet is cut, equal and opposite in strength and each equal and opposite to the original pole still attached to it. This is very much like the modern view, except that we now view elementary magnetism as a property of matter itself, not a separate fluid.

Wo have no proof that isolated magnetic poles do not exist somewhere, and there are no known reasons why they should not exist. Nevertheless the mathematical description of magnetism as we find it should reflect the empirical fact that no isolated magnetic pole has ever been detected. This description would have to be amended if isolated poles are ever discovered, but in the meantime it would include an essential fact of magnetism as presently experienced.

The impossibility of separating poles can be stated by saying that in any volume cut off physically from remaining space by a bounding surface, there is no net pole strength; in cutting through a magnet, you create a pole equal and opposite to one you were trying to surround. Now in electrostatics we are able to write a simple relation between the electric field intensity and its sources within a particular volume in terms of the *flux* of the field intensity. We can, similarly, define the flux of the magnetic field intensity $\vec{H}$ through an element of surface $\Delta\vec{S}$ as $\vec{H} \cdot \Delta\vec{S}$, and find that mathematically, as a result of the inverse square law, the total outward flux of $\vec{H}$ is

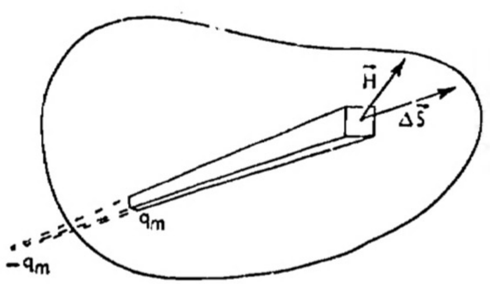


Figure 5.9: Finding total outward flux of $\vec{H}$.

$$\sum_{\text{s closed}} \vec{H} \cdot \Delta\vec{S} = 4\pi k' q_m$$

where q_m is the total polo strength inside the volume, as in Figure 5.9. (We recall that in the electrical case, owing to tho inverse square of the distance in tho quantity to be summed, it does not matter whore the charge is located within the volume, and the principle

of superposition then insures that tile total flux is independent of the distrihution of charge within the volume,) But the surface involved in tliis theorem of Gauss's is mathemati cal; there is no more physics in the new statement than in the original law of Michell and Coulomb.

The power of the Gauss form of Coulomb's law in electrostatics is that net electrical charges can be isolated with empty space completely surrounding them, that for any electric charge $+q$ there is no uniquely associated $-q$. As a result, Gaussian surfaces can be chosen with the same symmetry as that of the charge, so as to yield an expression for the electric field intensity. The analogous theorem for the flux of $\vec{H}$ is not so useful, since isolated spheres and lines of magnetic pole strength cannot be constructed.

And yet the concept of magnetic flux suggests a way of stating the in separability of poles *and* the inverse square law at the same time. Let us assume a quantity $\vec{B}$ which is indistinguishable from the magnetic field intensity in empty space *outside* the magnet but defined inside the magnet by the condition that its total outward flux from *every* closed surface is zero whether there arc magnets or not. Thus we may write

$$\sum_{\text{s closed}} \vec{B} \cdot \Delta\vec{S} = 0$$

for all possible surfaces. This condition is satisfied by $\vec{H}$ itself for surfaces which do not cut through magnets, but not in general for surfaces which do. The demand on $\vec{B}$ is equivalent to demanding a physical cut through the magnet, which would create a pole strength to cancel that already inside the volume, rather than the mere mathematical surface of Gauss's theorem for the field intensity. The behavior of lines of $\vec{B}$ through a magnet is shown in Figure 5.10.

We shall see in the following chapters that $\vec{B}$ can be given an operational definition in connection with another aspect of magnetism. The only virtue in introducing it here is to state in mathematical form the inseparability of poles as we find them in nature. In the MKS system of units, $\vec{B}$ and $\vec{H}$ are expressed in different units:

$$\vec{B} = \mu_0 \vec{H}$$

in empty space, where μ_0 is $4\pi \times 10^{-7}$. Again in empty space outside the magnet, a single pole of strength q_m webers gives rise to $\vec{B}$,

$$\vec{B} = \frac{q_m}{4\pi r^2}\vec{\hat{r}}$$

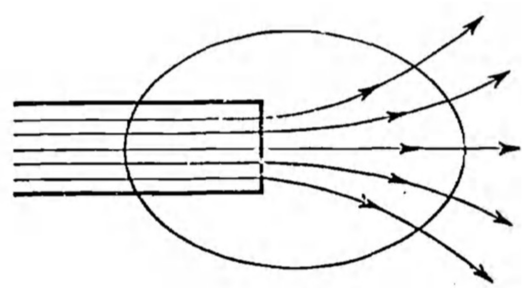


Figure 5.10: The behavior of lines of $\vec{B}$ through a magnet.

so that $\vec{B}$ is measured in webers per square meter. We should note the occurrence of the geometrical factor 4π, just as in electrostatics. One place or another this factor is sure to enter the description.

If this were all there were to magnetism, there would be no connection with electricity, and the two would be considered as separate subjects. It was, in fact, one of Gilbert's achievements that he distinguished clearly between magnetism, as produced by magnetite and magnetized iron, and static electricity as produced by rubbing glass with silk. But there is current electricity which consists of a net flow of charge, whether as free charge or in electrical conductors, and magnetism *is* connected with the motion of charge, as we shall see in the next chapter.

Problems

5.1 Suppose you were confronted with two iron bars that look identical in every respect, but one has been "permanently" and strongly magnetized with its poles well localized at the two ends, and the other not. Without any additional equipment whatsoever, how could you determine which is a permanent magnet and which has no residual magnetism of its own. Describe in detail the operations you could perform and the conclusions you would draw at each step.

5.2 When one end of a magnet is brought close to one end of an initially magnetized nail, the nail itself becomes a magnet and will attract other nails. The effect is even more pronounced if the end of the magnet is actually brought into contact with the end of the nail. (If you have not observed such phenomena you can easily do so with toy magnets obtained at a variety store.) What electrical phenomenon is this analogous to? Is the analogy complete, i.e.,, in what way do the phenomena differ? (Hint: What would happen in the electrical case after the two objects came in contact with each other?)

5.3 Make a detailed list of ways in which electrostatic and magnetostatic phenomena clearly differ from each other. (Note such aspects us the nature of materials that exhibit relevant properties, phenomena of conduction, phenomena of polarity, etc.)

Chapter 6

The Magnetic Interactions Of Steady Currents

The electrical effects first studied were those of static charges produced by rubbing amber with cloth, but no quantitative results wore obtained until the properties of conductors were distinguished from those of insulators. In electro*statics* a conductor is an object whose surface has an equilibrium distribution of charge: There is no difference of potential between any two points on the surface of a conductor. This follows from the definition of a conductor as something in which charge is free to move.

But if a difference of potential can be maintained between two points of a conductor, there will be a flow of charge from one point to the other. We no longer have a *static* situation, but we may have a *steady* flow. Let us consider a linear conductor such as a straight wire, and assume that some external device can maintain a constant difference of potential between the ends (see Figure 6.1). This device will need to supply charge, but the charge does not build up anywhere – it leaves the wire at the same rate that it enters, and the conductor need have no net charge. The result is a steady flow of charge in the wire. The amount of charge per unit time which passes any position P is the current:

Figure 6.1: Current flows I in a conductor due to potential difference V provided by a battery.

$$I = \frac{\Delta q}{\Delta t} \frac{\text{coulombs}}{\text{second}}$$

and one coulomb/second is called an ampere. The direction of the current is that of *positive* charge flow. A flow of negative charge is equivalent to a current whose direction is opposite to the motion of the charge.

Many practical uses of conductors involving the transfer of charge are treated in the laboratory monograph of this series. Here we shall not be concerned primarily with any particular relation between the magnitude of current in a conductor and the applied potential difference, but demonstration of the effect under immediate consideration does involve setting up electrical circuits, for which it usually suffices to know Ohm's law: For many conductors the current I of Figure 6.1 is directly proportional to the applied potential difference denoted by V in the figure. Mathematically,

$$V = IR$$

where R, called the resistance, depends on the material of which the conductor is made – whether copper or aluminum, for example. For a uniform wire the resistance is directly proportional to the length and inversely proportional to the cross-sectional area of the wire. Ohm's law is important as describing the behavior of many conductors, but it is not a fundamental law of electricity and magnetism. The magnetic effects of electric currents do not depend on the applicability of Ohm's law.

The first currents observed were those obtained by discharging conductors which had been charged electro statically, but such currents are usually small and sporadic. Production of fairly large steady currents became possible only after Volta's invention of the chemical battery at the beginning of the nineteenth century. Development of tho battery as a practical device facilitated many kinds of electrical experiments including tho effects of current electricity.

Even earlier (1752), Benjamin Franklin had demonstrated that lightning is an electrical discharge. Observation of occasional erratic behavior of compass needles during a thunder storm suggested to Hans Christian Oersted of Denmark some connection be tween electricity and magnetism, and led to his remarkable discovery in 1819 that current electricity is accompanied by magnetic effects. Oersted's experiment consisted of setting a long straight portion of an electric circuit above and parallel to a compass needle, and finding that the needle is deflected from its original north-south orientation when the circuit is closed (see Figure 6.2). This is not a temporary effect: The deflection is maintained so long as the current is maintained. With a strong current the needle is very nearly at right angles to the line of the current, and the deflection of the needle is reversed when the current is reversed. If the compass is held above the wire instead of below, and the direction of the current is unchanged, the deflection of the needle is again reversed.

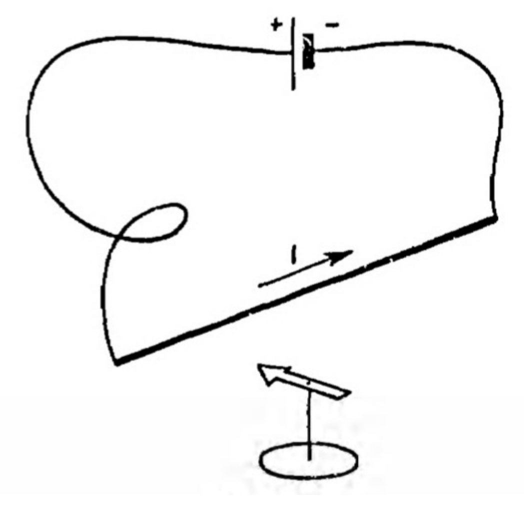


Figure 6.2: Magnetic effect of electric current. A current flowing in a wire will affect a magnetic needle nearby.

We have seen that a compass aligns itself along the direction of the magnetic field Intensity, and that field lines can be traced out with a small compass. The "sense" or direction of the arrow on a field line is that of the dipole moment of the compass: from negative to positive, or from S pole to N pole. The field lines so traced for a long straight wire carrying a current are circles, directed in accord with a *right-hand rule*: If the wire is grasped with the right hand, the thumb pointing in the direction of positive charge flow, the direction of the curved fingers is that of the field lines. (Check this rule with Figure 6.3.)

A quantitative study of the magnetic field intensity accompanying a long straight linear current was under taken by Biot and Savart in Paris, immediately after hearing of Oersted's discovery. They found that the magnitude of the magnetic field intensity $\vec{H}$ at any point is directly proportional to the strength of the current, and inversely proportional to the shortest distance from the point to the wire. Quantitatively,

$$\vec{H} = \frac{I}{2\pi r} \tag{6.1}$$

for a long straight wire carrying current of magnitude I, in MKS units. The current is measured in amperes and r in meters. There is no explicit arbitrary constant (for a change!), because the ampere and the weber, which we have already encountered in Chapter 5 as a unit of magnetic pole strength, are so defined that the only factor in this formula is 2π, the ratio of the circumference of a circle to its radius. Clearly the magnetic field intensity $\vec{H}$ may be expressed in amperes per meter instead of newtons/weber, and it is usually so expressed in MKS units. In fact, $\vec{H}$ can be defined by this equation, with the magnitude of $\vec{H}$ at a distance of one meter from a long straight wire carrying a current of one ampere being equal to $(1/2\pi)$ amperes/meter. The equation as it stands, however, does not give the direction of the magnetic field intensity, and the right-hand rule must be kept in mind as well as the relation of magnitudes.

The form of the Biot-Savart law given above, together with the right-hand rule, suggests another way of putting the relation between I and the magnitude of $\vec{H}$. Let us define what is called the circulation of $\vec{H}$: Consider a closed path, s, in the field; for every part of the path multiply the element of length Δs by the component of $\vec{H}$ parallel to Δs, and sum the product over the entire path. The result is called the circulation of $\vec{H}$ about the path chosen. For a circular path in a plane perpendicular to the current whose line

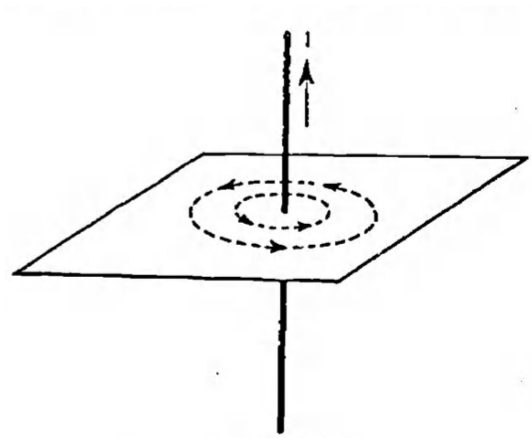


Figure 6.3: A current flowing in a wire will produce field lines around it. The direction of field lines is determined by the right-hand rule.

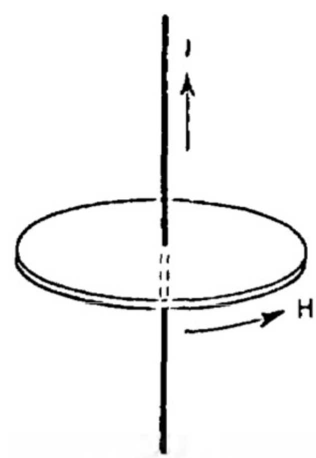


Figure 6.4: Circulation of $\vec{H}$ about a closed path is the current I.

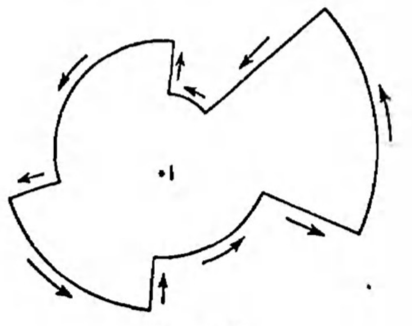


Figure 6.5: Finding circulation of $\vec{H}$ for a complicated path.

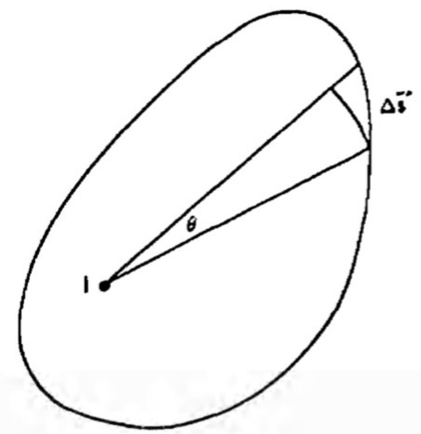


Figure 6.6: Contribution of a slant elements to circulation of $\vec{H}$ about a closed path depends on the angle subtended *theta*.

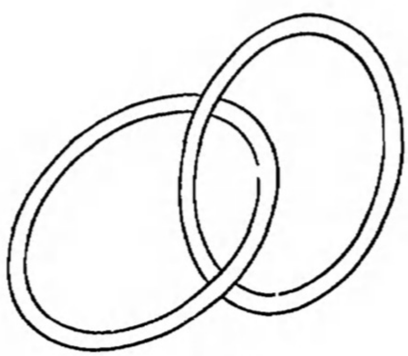


Figure 6.7: Shape of circuit does not matter for calculating the circulation of $\vec{H}$ about a closed path.

passes through the center of the circle (Figure 6.4), this process is easily carried out. The field intensity is everywhere in the same direction as the path, the entire path length is $s = 2\pi r$, and the circulation is simply

$$\sum_{s \text{ closed}} \vec{H} \cdot \Delta s = 2\pi r \vec{H} = 2\pi r \frac{I}{2\pi r} = I \tag{6.2}$$

just the current in the wire.

But let us evaluate the circulation of $\vec{H}$ over a somewhat more complicated path consisting of concentric circular arcs connected with radial lines as in Figure 6.5. The length of each arc is proportional to both its radius and the angle it subtends, but the field intensity is inversely proportional to the radius. The radial portions of the path con tribute nothing, since they are perpendicular to $\vec{H}$, so that the circulation about this path is I, just as before. Even a slant element of path contributes to the circulation only $\theta/2\pi$, where θ is the angle it subtends as shown in Figure 6.5. Thus the if circulation of $\vec{H}$ is equal to I if the area bounded by the closed path has I passing through it, and equals zero if the current circuit does not link through the loop. Otherwise, the shape of the closed path does not matter, nor does the exact position of tho current; the only consideration is whether the current "threads" the loop, that is, whether there is flow of charge through the area bounded by the loop (Figure 6.7).

That this result is perfectly general follows from its independence of tho shape of the loop for a line current, and from tho principle of superposition for $\vec{H}$. We may write

$$\text{circulation of} \quad \vec{H} = \sum_{s \text{ closed}} \vec{H} \cdot \Delta s = I \tag{6.3}$$

where I is the total (net) current threading the path of the circulation. Tho circulation of $\vec{H}$ about two equal and opposite currents is zero, even though the currents are displaced from each other and the magnetic field intensity itself may have quite appreciable values at various points along the path.

The circulation law for $\vec{H}$ is extremely useful in finding the field intensity associated with all current configurations which have cylindrical symmetry. An important example is the solenoid, a coil of insulated wire wound in a close helix or spiral on a hollow cylinder, or having the shape of of a cylinder. Let us consider a very long coil of this kind, of n turns per unit length, each carrying

current I. We may investigate the field intensity well away from both ends by taking a circulation path partly inside and partly outside the coil, as in Figure 6.8, barely including the current carrying wires. For each turn of wire the magnetic field in the plane of the turn is at right angles to the plane, and If the contributions of the loops above and below are considered in pairs it can be seen that the whole field is parallel to the axis of the solenoid, so long as we stay far from the ends.

Thus the short sides of the circulation rectangle contribute nothing, and

$$H_{\text{inside}} \times l - H_{\text{outside}} \times l = nlI$$

where l is the long dimension of the circulation rectangle. But this same equation holds if the circulation path is changed to the dotted line inside the coil, or for any other position of the inside leg of the rectangle. We can therefore conclude that the field intensity *inside* the cylinder is uniform, having the same strength and direction over the entire cross section. Exactly the same argument can be made for the leg of the rectangle outside the cylinder; the sides which are short in Figure 6.8 may be made as long as we please. The field intensity outside the solenoid, in a plane that cuts the cylinder far from the ends, is also uniform. But the field intensity outside must be very small indeed full justification of this statement is left to a problem. Therefore:

$$H_{\text{inside}} - H_{\text{outside}} = nI = H_{\text{inside}} \tag{6.4}$$

to a very good approximation.

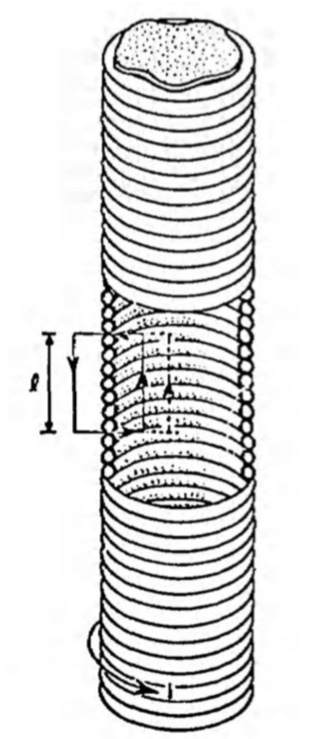


Figure 6.8: Investigating field intensity of a coil.

Except in current configurations of cylindrical symmetry, the circulation law is not so very useful in finding the magnetic field intensity accompanying a current. To express the field intensity at a point in terms of the current in a circuit of arbitrary geometry we shall need the vector product of two other vector quantities.

The prototype *vector* is almost literally the directed line segment in three-dimensional space by which other vector quantities such as force and velocity are represented. We are familiar with the scalar product of two vectors as it occurs in $\vec{F} \cdot \vec{s} =$ work, or $\vec{E} \cdot \vec{s} =$ electrical potential difference, giving a scalar quantity which can be expressed as a single number. The prototype cross product of two vectors is the area of the parallelogram defined by two directed line segments, represented in a direction perpendicular to that area. That two lines specify a definite parallelogram in a plane is shown

in Figure 6.9. By definition

$$\vec{C} = \vec{A} \times \vec{B} = AB \sin\theta\, \vec{c}$$

where $\vec{c}$ is a unit vector at right angles to the plane of $\vec{A}$ and $\vec{B}$. The sense or sign of $\vec{A} \times \vec{B}$ is determined by a right-hand rule: With the plane of your hand at right angles to the plane of $\vec{A}$ and $\vec{B}$, lot the fingers of your open right hand point in the direction of the first named vector ($\vec{A}$) with the hand oriented so that the parallelogram is in front of your palm – partial closing of the fingers would bring them parallel to the second vector ($\vec{B}$); the direction of your extended thumb is that of $\vec{A} \times \vec{B}$.

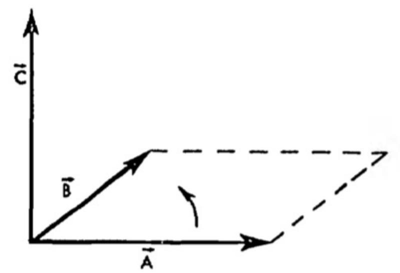


Figure 6.9: Cross multiplication of vectors.

Vector cross multiplication is not commutative: It is readily seen that

$$\vec{A} \times \vec{B} \neq \vec{B} \times \vec{A}$$

The cross product of $\vec{A}$ and $\vec{B}$ vanishes if $\vec{A}$ and $\vec{B}$ are parallel, and has its maximum magnitude if they are at right angles to each other. We note that an area is represented as normal to its plane, but that the sign of the normal is chosen by an arbitrary rule. Many physical quantities share this characteristic. Cross products may be represented by directed line segments, and, for most purposes, such as addition and multiplication, they behave like ordinary vectors. Actually a vector product is not exactly the same kind of quantity as at least one of its vector factors – that a directed area is not quite like a directed line segment is shown in one of the problems. In the problems the cross product is also expressed in terms of the components of $\vec{A}$ and $\vec{B}$ in Cartesian coordinates.

From other experiments of Biot and Savart and of Ampére (to whose further work we shall turn our attention shortly) on circular circuits and those which combine circular arcs and radial circuit elements, it became evident that magnetic effects are always directly proportional to the magnitude of the current, and that the effect of each element of current at a particular point P depends not only on the distance of the point, but also on the orientation of the current with respect to the line between it and the point at which $\vec{H}$ is to be determined. With reference to Figure 6.10, the contribution of a length of circuit Δs to the field intensity at a point whose distance is r is given by

$$\Delta H = \frac{I \Delta s \sin\theta}{4\pi r^2} \tag{6.5}$$

where θ is the angle between $I\Delta s$ (taken positive in the direction of the current) and the line between the current element and the

point. The direction of $\vec{H}$ is at right angles to both $I\Delta s$ and r, and in this instance into the page.

All this information is conveyed more simply by means of the cross product

$$\Delta\vec{H} = \frac{I\Delta s \times \vec{\hat{r}}}{4\pi r^2} \tag{6.6}$$

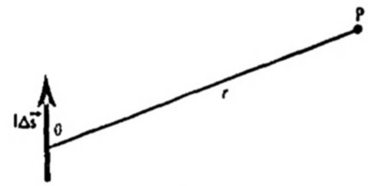


Figure 6.10: Finding the field strength.

where $\vec{\hat{r}}$ is a unit vector along r, directed *from* the current element *to* the point in space. This formula has to be inferred from experiments with complete circuits, for which the field intensity is correctly given by

$$\vec{H} = \sum_{s \text{ closed}} \frac{I\Delta s \times \vec{\hat{r}}}{4\pi r^2} \tag{6.7}$$

Let us apply this last formula to find the field intensity at the center of a circular loop of radius r, carrying current I. The lead wires from the battery produce no effect. (Why?) Since Δs is perpendicular to the radius of the loop, and the distance r is the same for all elements of current, the sum over all parts of the circuit can be evaluated at once:

$$H = \frac{I 2\pi r}{4\pi r^2} = \frac{I}{2r} \tag{6.8}$$

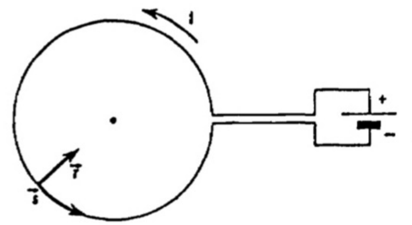


Figure 6.11: Finding the field strength for the given loop.

at the center of a circular loop. The direction of $\vec{H}$ is that of $\Delta\vec{s} \times \vec{\hat{r}}$, out of the page for the current indicated in Figure 6.11.

Ampére, who learned of Oersted's discovery at the same time as did Biot and Savart, reasoned that there should be forces between two current circuits if both produce magnetic effects, since two magnets interact with each other. Within a week ho had shown that two parallel wires carrying currents in the *same* direction (see Figure 6.12) attract each other, and repel each other if the currents are in *opposite* directions. The magnitude of the force between the wires is directly proportional to both currents.

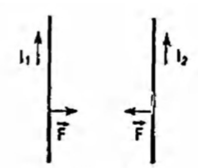


Figure 6.12: Finding force between two parallel wires carrying current.

As the result of a remarkable series of experiments performed during the next three years, Ampére was able to infer that the force between two parallel current elements $I_1\,\Delta s_1$, and $I_2\,\Delta s_2$ is given by

$$\begin{aligned}\Delta F &= \frac{k' I_1 I_2 \sin\theta \Delta s_1 \Delta s_2}{r^2} \\ &= \frac{\mu_0}{4\pi}\frac{I_1 I_2 \Delta s_1 \Delta s_2 \sin\theta}{r^2}\end{aligned}$$

where $\mu_0/4\pi$ is the constant of proportionality in MKS units (see Figure 6.13). The value of k' is taken arbitrarily as 10^{-7} N/A^2, so

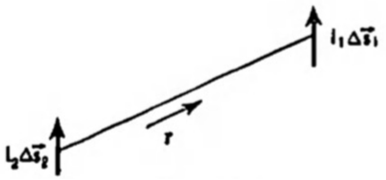


Figure 6.13: Finding force between two parallel wires carrying current.

that $\mu_0 = 4\pi \times 10^{-7}$ N/A^2. The size of the ampere (and thus also the coulomb) is in fact determined by taking the constant of exactly this magnitude.

Again the presence of $\sin\theta$ suggests a cross product. Still another angle must be taken into account if the two current elements are not parallel, and the formula begins to look even more complicated:

$$\begin{aligned} \Delta\vec{F}(\text{on } I_1\Delta\vec{s}_1) &= \frac{\mu_0}{4\pi}\frac{I_1\Delta\vec{s}_1 \times (I_2\Delta\vec{s}_2 \times \hat{r})}{r^2} \\ &= I_1\Delta\vec{s}_1 \times \left(\frac{\mu_0}{4\pi}\frac{I_2\Delta\vec{s}_2 \times \hat{r}}{r^2}\right) \end{aligned} \tag{6.9}$$

where $\hat{r}$ is a unit vector directed from $\Delta\vec{s}_2$ toward $\Delta\vec{s}_1$. For the total force on current element $I_1\Delta\vec{s}_1$ exerted by circuit 2, we must sum over all $I_2\Delta\vec{s}_2$:

$$F(\text{on } I_1\Delta\vec{s}_1) = I_1\Delta\vec{s}_1 \times \left(\frac{\mu_0}{4\pi}\sum_{s_2 \text{ closed}}\frac{I_2\Delta\vec{s}_2 \times \hat{r}}{r^2}\right)$$

But this way of writing tho force suggests a simplification, since tho term in parentheses is, apart from the constant μ_0, the magnetic field intensity found by Biot and Savart as a force per unit pole. Moreover, it is found that a current element experiences a force when placed in the vicinity of an ordinary magnet, as indeed we might expect from the fact that a current exerts forces on a magnetic compass needle. The force on our current element $I_1\Delta\vec{s}_1$ may be written as

$$\vec{F} = I_1\Delta\vec{s}_1 \times \mu_0\vec{H} = I_1\Delta\vec{s}_1 \times \vec{B} \tag{6.10}$$

where

$$\vec{B} = \sum_{s_2 \text{ closed}} = \frac{\mu_0}{r\pi}\frac{I_2\Delta\vec{s}_2 \times \hat{r}}{r^2} \tag{6.11}$$

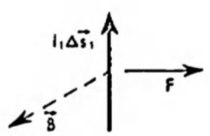


Figure 6.14: Finding force between two wires carrying current.

if $\vec{B}$ is produced by Current I_2. In general throughout empty space

$$\vec{B} = \mu_0\vec{H} \tag{6.12}$$

whatever the sources of magnetic field intensity (see Figure 6.14).

In Chapter 5 the field $\vec{B}$ was defined so that its net flux through the surface enclosing any volume of space vanishes, although the net flux of $\vec{H}$ might be different from zero. The lines of *both* $\vec{B}$ and $\vec{H}$ as produced by currents are without beginning or end, and

the net flux of both quantities through the surface enclosing any volume is zero. To see this we need consider only an increment of field arising from a single current element $I\Delta\vec{s}$, and remember that the entire field at a point is the sum of such increments. Let us examine

$$\Delta\vec{B} = \mu_0 \Delta\vec{H} = \frac{\mu_0 I}{4\pi}\frac{\Delta\vec{s}\times\hat{r}}{r^2}$$

The direction $\Delta\vec{B}$ is perpendicular to both Δs and $\hat{r}$ and its magnitude depends on r. If we choose any point and move along the direction of $\Delta\vec{B}$, we shall trace out a circle in a piano at right mimics to the direction of $\Delta\vec{s}$, which would yield no not flux from tho surface of a volume such as shown in Figure 6.15. Increments of $\vec{B}$ arising from other current elements have the same property: The circles of $\Delta\vec{B}$ they contribute may lie in different pianos, but the lines of each are continuous. Thus for $\vec{B}$ as a whole, the summation of the flux over any closed surface,

$$\sum_{s \text{ closed}} \vec{B}\cdot\Delta\vec{s} = 0 = \sum_{s \text{ closed}} \vec{H}\cdot\Delta\vec{s}$$

if $\vec{H}$ is produced by currents.

The equation $\vec{F} = I\,\Delta\vec{s}\times\vec{B}$ is often taken as the definition of the vector field $\vec{B}$, which is called the magnetic induction field. The units of $\vec{B}$ are newtons/ampere-meter. It is left to the problems to show that these units are consistent with those given in Chapter 5. In empty space, where our equations hold, the field quantities $\vec{B}$ and $\vec{H}$ are really indistinguishable, although their units are arbitrarily different in the MKS system. We shall investigate further the equivalence of currents and magnets, and how this equivalence depends on the absence of separable magnetic poles, but let us first look again at the role of $\vec{B}$ in the interaction of two currents.

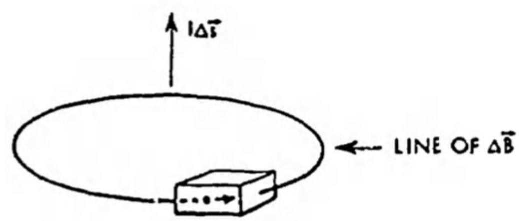


Figure 6.15: Finding force between two parallel wires carrying current.

We have investigated the magnitude and direction of $\vec{B}$ (although we some times called it $\vec{H}$) in relation to its sources in some simple cases. If we know $\vec{B}$ at any point we can immediately find the force on a current element $I\Delta\vec{s}$ placed at that point by computing $I\Delta\vec{s}\times\vec{B}$. In terms of Ampére's experiments, the interaction between two currents is thus for convenience considered in two steps, the production of a field $\vec{B}$ by one circuit and the action of the field $\vec{B}$ on the other circuit.

In view of the complicated dependence of the forces on the angles involved, this procedure has great advantages, since we need

now consider only one angle at a time, furthermore, the contributions to $\vec{B}$ from different sources, be they current or magnets, are additive, so that simultaneous interactions may be considered in a relatively simple way. Even so, it should be remarked that the greatest advantages of the field concept become apparent only when tho sources, and therefore the fields, are permitted to vary in time. The subject of time varying fields is reserved for another monograph in this series, but we should note that light and other electromagnetic radiation can be simply understood only on the basis of electric and magnetic field quantities.

If we consider only forces on current elements, we need only one magnetic field quantity, that defined to give the force per unit current at right angles to the direction of the field, namely, $\vec{B}$. The necessity for considering a second field quantity, the magnetic field intensity $\vec{H}$, does not then arise until we consider magnetic materials, within which $\vec{B}$ and $\vec{H}$ are different. But *can* we confine our attention exclusively to currents? It was Ampére's hypothesis that *all* magnetic interactions can in fact be traced to currents, whether they occur in macroscopic circuits or are assumed to exist in the most elementary form of matter. To establish the basis for this hypothesis we must consider the forces on a loop of current in a field $\vec{B}$.

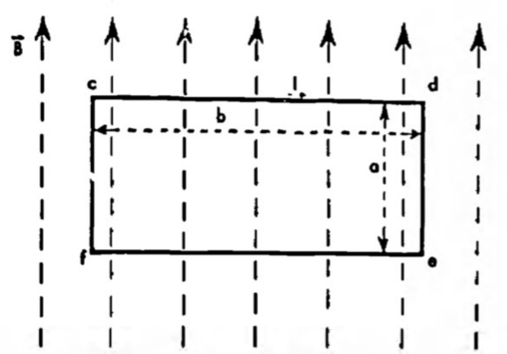


Figure 6.16: Forces on a loop of current in a field $\vec{B}$.

Let us take a rectangular loop of wire edef, as shown in Figure 6.16, having dimensions a and b, placed initially so that its plane is parallel to a field $\vec{B}$ which is uniform in space and constant in time. The current in the loop is I. We nay compute the force on each straight section of the loop from the formula $\vec{F} = I\Delta\vec{s} \times \vec{B}$. With the current as Indicated, we see that there is a force IbB directed *out* of the plane on the wire cd, a force IbB directed *into* the plane on the wire ef and no forces on fc and de since these wires are in the sane direction as the lines of $\vec{B}$. The net force on the loop is zero, but there is a torque of magnitude $abIB$ tending to turn it out of the plane, into such a position that what is now the front of the loop faces down, with the plane of the loop perpendicular to the lines of $\vec{B}$. This is exactly what would happen to a rectangle of magnetic material whose front face is a negative (south-seeking) pole and whose back face is positive (north-seeking). The loop is thus equivalent to what Ampere called a magnetic shell, a flat magnet of magnetic moment proportional to the product of its area and the current on its boundary.

We may recall that the torque on a magnet of magnetic moment

$\vec{m}$ in a region of field intensity $\vec{H}$ is mH when the direction of the moment is at right angles to $\vec{H}$. The torque on our loop is IAB, where $A = a \times b$ is the area of the loop. If we want to keep the same units as before for magnetic moment, we may ascribe to the loop a moment of magnitude $\mu_0 IA$, directed perpendicular to the plane of the loop, and positive toward a right-hand thumb whose curved fingers point along the current. For other orientations of the loop the torque is $IAB \sin\theta$, where θ is the angle between the magnetic moment and the field lines, in agreement with the torque $mH \sin\theta$ on a magnet in a field intensity $\vec{H}$.

The magnetic moment of a current does not depend on the shape of the loop. A circular loop is equivalent to a magnetic disk whose faces are of equal and opposite polarity, and whose magnetic moment is again $\mu_0 IA$, with A the area of either face. Ampére went further; according to his hypothesis, all magnets are current configurations, which exist on a submicroscopic scale.

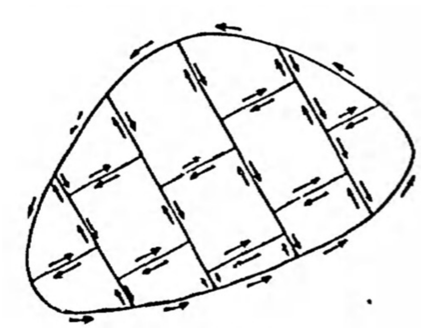


Figure 6.17: A "magnetic shell" has zero net current except at boundary.

A "magnetic shell" would consist of an indefinite number of tiny current whirls, all oriented in the sane sense, so that the net current is zero except at the boundary of the shell, or loop; internally the currents of contiguous whirls cancel each other as is evident in Figure 6.17. A helix of wire with adjacent turns of current is thus equivalent to a stack of magnetic disks, as in Figure 6.18; the net effect as a magnet is a positive (north-seeking) pole at one end of the helix and a negative pole at the other. The impossibility of separating pules is then just the impossibility of separating the faces of a disk.

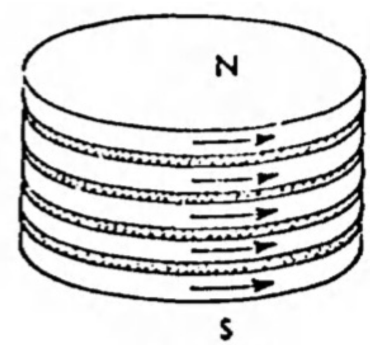


Figure 6.18: A stack of magnetic disks is equivalent to a helix of wire with adjacent turns of current.

The modern view of magnetism is not very different from this simple picture, and Ampére's hypothesis has been accepted in principle. Because of the absence of separable magnetic poles, all magnetism is traced to currents, even when the physical currents are not accessible to measurement. The neutron, for example, is an uncharged particle, but it does have a magnetic moment; in this sense it behaves like a circulating negative charge. Recent experiments have shown that the neutron does behave more like an infinitesimal current whirl than like an infinitesimal linear magnet, although its current is quite inaccessible for detailed investigation.

Search for the isolated magnetic pole continues, but all the evidence in hand supports the view that $\vec{B}$, a field which acts on currents and can be traced to moving charge, is a more basic concept than that of the magnetic field intensity $\vec{H}$, simply because electric charges and currents do exist, nevertheless we continue to explore magnetic

fields with iron filings and compass needles, and the concept of magnetic field intensity $\vec{H}$ can hardly be avoided in the description of magnetic materials.

Einstein has pointed out in his "scientific autobiography" that concepts used to describe the physical world are in truth intellectual *creations*, devised by scientific imagination but not freely: The discipline imposed by observation and experiment is very strict. The concepts of electromagnetic theory are very intimately related to one another, reflecting an enormous body of diverse but related experimental results. More than one of the concepts plays a dual role, and experiments can often be described in more than one way. Yet we shall see in Monograph III that there are only four fundamental empirical laws of electricity and magnetism, which can be described In a must elegant and simple way in terms of field quantities. We have already considered three of these laws: Coulomb's law for the interaction of two static charges, the law of Michell and Coulomb for the interaction of two magnetic poles plus a statement of the inseparability of poles, and the law describing the magnetic effect of currents or the interaction of two currents, usually called Ampére's law. The remainder of this booklet will be devoted to further development of this third law. Before undertaking a more mathematical and thus more powerful formulation of the law, however, let us examine the effect of magnetic fields on individual charges.

The currents we have considered thus far were assumed to exist in uncharged conductors, but a current is defined as a flow of charge. The question of whether the mechanical motion of a charged body produces magnetic effects was first tested experimentally by Henry Rowland of Johns Hopkins University in 1878. Rowland electrostatically charged the rim of an insulating disk, rotated the disk, and found that magnetic effects were indeed produced. The experiment is very difficult perform because the currents produced in this way are small, but the result is unambiguous. That moving charges experience a force in a magnetic field in much easier to demonstrate: Streams of electrons in a cathode ray lube produce a visible glow on the glass envelope of the tube which is shifted very readily by even a small magnet. In fact, it was concluded that cathode rays are charged particles as a result of their deflection by *both* electric fields and magnetic fields.

The correct formula for the force on a charge moving in a magnetic field can be obtained from that for the force on an element of current in a conductor. Let us assume that a conductor has N movable charges per unit volume, each charge of magnitude q.

These movable charges may undergo very complicated motions, but if there is a net flow of charge in one direction, we may ascribe to them an average "drift" velocity v. A linear conductor of *unit* cross section, a segment of which is indicated in Figure ??, would then carry a current Nqv – this is the amount of charge crossing the face shown per unit time. The amount of charge per unit time flowing through a portion of the face whose area is A is then $NqAv$, and the general formula is $I = NqAv$, For a current element of length Δs,

$$I\,\Delta s = NqA\Delta s\,v$$

But $A\Delta s$ is just the volume of the current element and $NA\Delta s$ is the total number of charges involved. Thus the force per charge q is

$$\vec{F} = \frac{I\,\Delta s \times \vec{B}}{NA\Delta s} = q\,\vec{v} \times \vec{B} \tag{6.13}$$

This derivation of $\vec{F} = q\,\vec{v} \times \vec{B}$ involves too many assumptions to be rigorous, but the result is entirely correct. For a charge q moving with velocity $\vec{v}$, $q\vec{v}$ is equivalent to a current element. The force $q\,\vec{v} \times \vec{B}$ is called the Lorentz force, first derived rigorously by the famous Dutch physicist H.A. Lorentz in 1892. It is often taken as a fundamental equation, and the force on a current element derived from it. It can be taken as the defining equation for the magnetic field quantity $\vec{B}$: $\vec{B}$ is that field which gives a *velocity dependent* force on a charge q, in accord with the equation for the Lorentz force, as distinguished from the electric field intensity $\vec{E}$, which produces a force which is independent of the velocity. We note that both $\vec{E}$ and B are defined in a particular frame of reference, that in which the velocity of the charge is $\vec{v}$.

One of the most interesting properties of the Lorentz force is that it is incapable of changing the speed or kinetic energy of a moving charge, since the force (and therefore the acceleration) is at right angles to the velocity. In a uniform magnetic field which is itself perpendicular to the velocity, the motion of a charged particle is circular.

Problems

6.1 Show that if $\vec{\hat{x}}$, $\vec{\hat{y}}$ and $\vec{\hat{z}}$ are unit vectors in the direction of increasing x, y, and z in a right-handed Cartesian coordinate

system,

$$\begin{aligned}
\hat{x}\times\hat{x} &= 0 = \hat{y}\times\hat{y} = \hat{z}\times\hat{z}\\
\hat{x}\times\hat{y} &= 0 = \hat{z}\\
\hat{y}\times\hat{z} &= 0 = \hat{x}\\
\hat{z}\times\hat{x} &= 0 = \hat{y}
\end{aligned}$$

Show also that in terms of Cartesian component a the vector product of $\vec{A} = A_x\hat{x} + A_y\hat{y} + A_z\hat{z}$ and $\vec{B} = B_x\hat{x} + B_y\hat{y} + B_z\hat{z}$ may be written in the form of a determinant:

$$\vec{A}\times\vec{B} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z}\\ A_x & A_y & A_z\\ B_x & B_y & B_z \end{vmatrix}$$

6.2 Suppose you have a long cylindrical conductor of radius r_0 carrying total current I_0. The current is distributed uniformly over a cross section of the conductor, so that $I_0 = \pi r_0^2 j$, where j is the current density in amperes/meter^2.

(a) The field lines of $\vec{H}$ outside the cylinder arc circles centered at the axis. Would this also be true of the magnetic field lines within the cylinder? Give your reasons .

(b) What is the magnetic field intensity on the axis of the cylinder?

(c) What is the magnitude of $\vec{H}$ at $r = \frac{1}{3}r_0$

(d) Plot the magnitude of $\vec{H}$ against the distance r from the axis of the cylinder, both inside and outside the cylinder. At what distance from the axis is this magnitude greatest?

6.3 Let P be a point outside the long solenoid of Figure 6.8, whose distance from the axis is large compared with the radius of the solenoid. According to (6.6), each element of current contributes to the field intensity $\Delta\vec{H} = I\Delta\vec{s}\times\frac{\vec{r}}{r^2}$

(a) Consider *qualitatively* the contributions to the magnetic field intensity from current elements $I\Delta s$ parallel to $\vec{r}$ and those perpendicular to $\vec{r}$ from a single turn of the coil in the plane of P perpendicular to the axis. Do they tend to reinforce or to cancel each other?

(b) Make the same qualitative estimate of contributions $\Delta\vec{H}$ for a point *inside* the solenoid.

(c) Consider, again qualitatively, the contributions $\Delta\vec{H}$ from two turns, one above and one below the plane of P perpendicular to the axis .

(d) What are your conclusions concerning the validity of (2.4)?

6.4 If a charged body moves with velocity $\vec{v}$ at right angles to a uniform field $\vec{B}$, its path is circular, since $\vec{F} = q\vec{v} \times \vec{B}$, and thus its acceleration is perpendicular to $\vec{v}$. What sort of path would result if $\vec{v}$ were in the same direction as the lines of $\vec{B}$? What sort of path would be produced if the angle between $\vec{v}$ and $\vec{B}$ were neither 0° nor 90° ?

Chapter 7

Magnetostatics Reformulated

Ampére's law is all there is to magnetostatics if we confine our attention to the magnetic effects of steady currents, but some aspects of the subject become apparent only if more elegant mathematical methods are available. Mathematics is more than a powerful tool for the solution of problems: Relations between physical quantities are often revealed by mathematical analysis, so that more physics emerges clearly. The danger lies in a tendency to substitute mathematical formalism for physical thought, to overlook or neglect the physical content of a mathematical equation or a line of mathematical reason, instead of being guided by it. Let us try to keep the physics, or sometimes only geometry, firmly in mind in our further analysis of magnetostatics, For this chapter we shall only assume knowledge of basic calculus, that branch of mathematics invented by Newton not only to make hard problems easier but also to sharpen and clarify his ideas concerning the physical world.

Let us begin by rewriting the relations considered earlier in terms of differentials and integrals. Ampére's law has appeared in essentially two different forms. The direct expression of the forces on a current element $I_1 \mathrm{d}\vec{s}_1$ is

$$\vec{F}\,(\text{on } I_1 \mathrm{d}\vec{s}\,) = I_1 \mathrm{d}\vec{s}_1 \times \vec{B} \tag{7.1}$$

where

$$\vec{B} = \frac{\mu_0}{4\pi} \oint \frac{I_1 \mathrm{d}\vec{s} \times \vec{\hat{r}}}{r^2} \tag{7.2}$$

where $\vec{\hat{r}}$ is a unit vector from $I_2\,\mathrm{d}\vec{s}_2$, to the point at which $\vec{B}$ is computed. The substitution of $\mathrm{d}\vec{s}$ for $\Delta\vec{s}$ is routine, but the circle at the middle of the integral sign is shorthand notation for the is closed below the summation sign of Chapter 6. Here it should be read

"the integral over the entire circuit s" of the integrand as written. This integral is, of course, a vector, every element contributing to $\vec{B}$ in a direction at right angles to both $\mathrm{d}\vec{s}_2$ and to the line from $\mathrm{d}\vec{s}_2$, to the position of $I_1 \mathrm{d}\vec{s}_1$ where $\vec{B}$ is to be determined to give the correct force.

We have also noted that $\vec{B}$ so defined yields a flux such that the total outward flux of $\vec{B}$ through any closed surface is zero. If $\mathrm{d}\vec{S}$ is an element of surface, directed *out* from the volume inclosed by the surface,

$$\int_{\text{closed}} \vec{B} \cdot \mathrm{d}\vec{S} = 0 \tag{7.3}$$

Unfortunately one must write in words that this surface integral is closed. A little sphere at the middle of the integral sign would be a convenient shorthand notation except for the fact that a sphere is indistinguishable from a circle in two dimensions. In the booklet on electrostatics we have considered a function corresponding to such a closed surface integral which describes a property of the integrand vector at each point in the inclosed space. We shall consider this property of $\vec{B}$ at a later stage of the discussion.

We have seen that the physical content of Ampére's law may be written alternatively as a circulation law for $\vec{B}$, which may also be written as an integral. The field $\vec{B}$ is a vector quantity defined at all points (x, y, z) in the region of interest. The *line integral* of $\vec{B}$ on a path C from point P to point P' is a scalar, written $\int_P^{P'} \vec{B} \cdot \mathrm{d}\vec{s}$, where $\mathrm{d}\vec{s}$ is an element of length in the direction of the local tangent to the path, and the integrand is the product of $\mathrm{d}\vec{s}$ and the component of $\vec{B}$ parallel to $\mathrm{d}\vec{s}$ (see Figure 7.1). (In terms of magnetic poles this integral would represent fig times the work done by the field in moving a unit pole from P to P'.) The circulation of a vector implies a *closed* path, not necessarily in a plane, and the integral sign can again be written with a circle at the middle to indicate that the initial and final points are coincident:

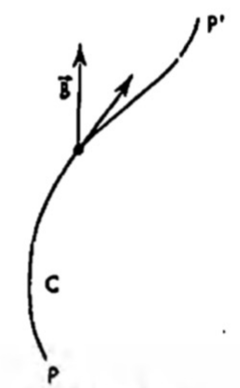


Figure 7.1: Component of $\vec{B}$ parallel to the path.

$$\oint_C \vec{B} \cdot \mathrm{d}\vec{S} = \mu_0 I \tag{7.4}$$

is simply a neater way of writing the circulation law for $\vec{B}$.

We have thus far considered almost exclusively linear currents such as those carried by thin wires, but the total current through a surface bounded by the circulation loop may be distributed over the surface and may vary from one part of the surface to the next.

We may readily take account of such variations if we express the total current I in terms of the current density $\vec{j}$. For a conductor of cross-sectional area A in which the density of current is uniform, $I = jA$. As a vector quantity, $\vec{j}$ is the current per unit area of a plane normal to the direction of charge flow. Since an element of surface can be represented by a vector $\mathrm{d}\vec{S}$ normal to its surface,

$$I = \int \vec{j} \cdot \mathrm{d}\vec{S} \tag{7.5}$$

is the total current through a surface over which the integral is evaluated (see Figure 7.2).

(In this monograph we confine our attention to steady currents, for which the integral of $\vec{j}$ over a closed surface would net zero.) Therefore the circulation law for E may be written

$$\oint_C \vec{B} \cdot \mathrm{d}\vec{s} = \mu_0 \int \vec{j} \cdot \mathrm{d}\vec{S} \tag{7.6}$$

where the integral on the right is to be carried out over a surface, in fact any surface, which is bounded by the curve C. We should recall that the positive direction for $\cdot\mathrm{d}\vec{S}$ is chosen by a right-hand rule; for a curve traversed counterclockwise in the plane, $\mathrm{d}\vec{S}$ is positive *out* of the page; if clockwise, $\mathrm{d}\vec{S}$ is positive *into* the page. Thus far we have hardly changed the form of the equations relating magnetic fields to their sources: $\vec{B}$ is written as an integral over all portions of a linear current, instead of the corresponding sum, and the line integral of $\vec{B}$ is related to current through a surface, as was the sum of $\vec{B} \cdot \Delta\vec{s}$ in Chapter 5.

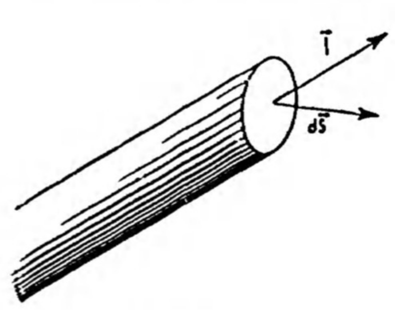


Figure 7.2: Vector normal to the surface.

Is it possible to relate the magnetic field to its source strength at each point, much as we found in the *Electrostatics* monograph that we could relate the electric field intensity to the charge density at each point? If there were only magnets with poles, and no magnetic effects of currents, the answer would be completely analogous to the electrostatic relations. It is indeed true that the net outward flux of the magnetic field intensity $\vec{H}$ (but not that of $\vec{B}$) from a closed volume is the total pole strength within the volume, and that $\vec{H}$ is related to density of pole strength in the same way that the electric field intensity is related to the charge density. But we have seen that the net flux of both $\vec{H}$ and $\vec{B}$ from a closed volume is zero if these fields are produced by currents. This is a very interesting and important property of magnetic fields, but it is of no immediate help in relating fields to the currents which produce them.

The answer comes from the circulation law for $\vec{B}$ (or for $\vec{H}$, since the two fields are the same in empty space, except for an arbitrary constant factor μ_0). It will be necessary to examine circulation more closely and to develop further the mathematics associated with it. This mathematics applies to all vector fields, but we shall call the field $\vec{B}$, and feel free to apply the mathematical consequences to other vector fields as the need arises.

In electrostatics we were guided to the appropriate mathematical theorem by Gauss's law: Beginning with a relation between the surface integral of the electric intensity $\vec{E}$ over a closed surface and the charge within the enclosed volume, we found that the function of $\vec{E}$ which can be identified with the charge density at each point in space is the divergence of $\vec{E}$ (div $\vec{E}$). The circuital form of Ampére's law relates a closed line integral of $\vec{B}$ to the integral of the current density over a surface bounded by the line. By analogy we should expect to find a function of $\vec{B}$ at each point of space such that its integral over any surface bounded by a line is equal to the circulation of $\vec{B}$ about the perimeter of the surface. If it is to be identified with the current density, it must be a vector quantity. In other words, we seek a vector $\vec{C}$ which is a mathematical function of $\vec{B}$ such that

$$\oint \vec{B} \cdot \mathrm{d}\vec{s} = \int \vec{C} \cdot \mathrm{d}\vec{S} = \mu_0 \int \vec{j} \cdot \mathrm{d}\vec{S} \tag{7.7}$$

If the second equality is to be true for any surface bounded by tho curve of of the line integral, it follows that the integrands must be the same, and $\vec{C} = \mu_0 \vec{j}$.

Consider any simple closed curve in a region where there is a field $\vec{B}$. An area bounded by a closed curve, whether plane or not, can be divided into two or several areas by lines between two points on the boundary. In Figure 7.3 there are three areas, around each of which we may take the circulation of $\vec{B}$ in the counterclockwise sense indicated. In the sum of these circulations all the Interior boundaries are traversed twice, in opposite directions. Thus

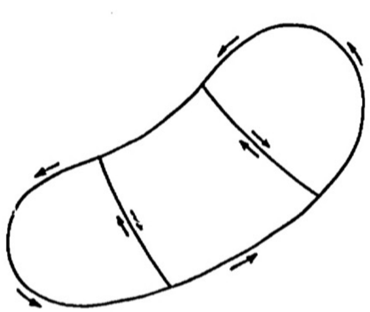


Figure 7.3: Taking circulation of $\vec{B}$ on a simple closed curve.

$$\oint \vec{B} \cdot \mathrm{d}\vec{s} = \oint \vec{B} \cdot \mathrm{d}\vec{s}_1 + \oint \vec{B} \cdot \mathrm{d}\vec{s}_2 + \oint \vec{B} \cdot \mathrm{d}\vec{s}_3 \tag{7.8}$$

The sum of counterclockwise circulations of $\vec{B}$ about all the subdivisions is just the circulation about the whole, and this result is independent of whether the surface, or the curve bounding it, is plane. The same is true of clockwise circulations, of course, but

the sense of all the circulations in the sum must be the sane for cancellation of adjacent interior boundaries. The rule is equally Justified for ten subdivisions, or a hundred. In general

$$\oint \vec{B} \cdot d\vec{s} = \sum_i \oint \vec{B} \cdot d\vec{s}_i \tag{7.9}$$

(We are here reminded of Ampére's hypothesis on the equivalence of a closed current and a shell of magnetic dipoles which involves submicroscopic current whirls; there is indeed a similarity, which we shall pursue later.)

Circulation itself is a scalar quantity, but any plane surface has an orientation in space which can be specified by a vector normal to the plane, and any well behaved surface can be broken up into elements sufficiently small to be considered plane. (We shall exclude surfaces with infinite peaks.) Let us see what the orientation of a surface has to do with the circulation of a vector about its boundary .

Consider a small triangular plane boundary *a b c*, Figure 7.4, and two surfaces bounded by it, one the plane *abc*, the other a surface made up of segments of three planes chosen at right angles to each other. According to the sum rule,

$$\oint_{abc} \vec{B} \cdot d\vec{s} = \oint_{adc} \vec{B} \cdot d\vec{s}_1 + \oint_{bcd} \vec{B} \cdot d\vec{s}_2 + \oint_{and} \vec{B} \cdot d\vec{s}_3 \tag{7.10}$$

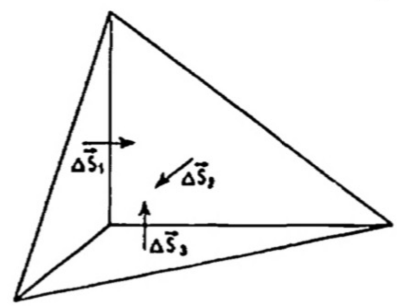


Figure 7.4: Taking circulation of a small triangular plane boundary.

It is most unlikely that the three terms on the right contribute equally to the total circulation, for several reasons. Each small circulation specifics a different area and a different length of perimeter, and of course $\vec{B}$ itself and its variation in space is quite independent of the surfaces we happened to choose. There is a relation between the surfaces themselves, which follows from the theorem that the vector sum of all the outward surfaces of a polyhedron equals zero. (The proof of this theorem is left to a problem.) We note that the four triangles bound a tetrahedron (Figure 7.4 and Figure 7.5), for which the right-hand rule for circulation leads to positive inward directions for surfaces 1, 2, and 3, and a positive outward surface for the triangle abc. If $\Delta\vec{S}_1$, $\Delta\vec{S}_2$, $\Delta\vec{S}_3$ and $\Delta\vec{S}$ represent these four surfaces, then

$$\Delta\vec{S} = \Delta\vec{S}_1 + \Delta\vec{S}_2 + \Delta\vec{S}_3 \tag{7.11}$$

But does this vector relation between surfaces have anything to do with the scalar relation of the circulations? If there is a vector related to the circulation about each surface such that its scalar product

with the surface itself would net the circulation the answer would certainly be affirmative .

A vector related to the circulation about an elementary plane surface can be constructed by multiplying the circulation of $\vec{B}$ about its boundary by a unit normal to the surface, and dividing by the magnitude of the surface:

Let us write $\oint \vec{B} \cdot \mathrm{d}\vec{s}_1, /\Delta S_1)\vec{\hat{n}}_1$ as a vector in the direction of the unit normal $\vec{\hat{n}}_1$, to $\Delta\vec{S}_1$. The scalar product of this vector with $\Delta\vec{S}$ is just $\oint \vec{B} \cdot \mathrm{d}\vec{s}_1$, since $\vec{\hat{n}}_1 \cdot \Delta\vec{S} = \Delta S_1$. In fact, if we take the vector

$$\vec{C} = \frac{\oint \vec{B} \cdot \mathrm{d}\vec{s}_1}{\Delta S_1}\vec{\hat{n}}_1 + \frac{\oint \vec{B} \cdot \mathrm{d}\vec{s}_2}{\Delta S_2}\vec{\hat{n}}_2 + \frac{\oint \vec{B} \cdot \mathrm{d}\vec{s}_3}{\Delta S_3}\vec{\hat{n}}_3 \tag{7.12}$$

then $\vec{C} \cdot \Delta\vec{S}$ is the original sum of circulations. We note that $\vec{C}$ is not in general the same as $\oint \vec{B} \cdot \mathrm{d}\vec{s}_1/\Delta S_1\, \vec{\hat{n}}_1$, where $\vec{\hat{n}}_1$ is the unit vector normal to the surface *abc*, although $\vec{C} \cdot \Delta\vec{S} = \oint \vec{B} \cdot \mathrm{d}\vec{s}$. The quantity $\oint \vec{B} \cdot \mathrm{d}\vec{s}/\Delta S$ is tho component of $\vec{C}$ normal to the surface $\Delta\vec{S}$.

The components of $\vec{C}$ are thus far defined in relation to plane surfaces, which nay be oriented in three mutually perpendicular and thus independent directions. Our surfaces $\Delta\vec{S}_1$, $\Delta\vec{S}_2$, and $\Delta\vec{S}_3$, are themselves components of $\Delta\vec{S}$. Now the surface $\Delta\vec{S}$ may be taken as small as we please, and in the limit of small $\Delta\vec{S}$ we define a vector which is called the curl of $\vec{B}$, written curl $\vec{B}$. For the component of curl $\vec{B}$ normal to a surface element ΔS whose orientation is $\vec{\hat{n}}$

curl $\vec{B}$ is also written sometimes as $\nabla \times \vec{B}$.

$$(\mathrm{curl}\,\vec{B})_{\vec{\hat{n}}} = \lim_{\Delta S \to 0} \frac{\oint \vec{B} \cdot \mathrm{d}\vec{s}}{\Delta S} \tag{7.13}$$

with the line integral taken around the boundary of ΔS. The integral form of the circulation sum is then

$$\oint \vec{B} \cdot \mathrm{d}\vec{s} = \int \mathrm{curl}\,\vec{B} \cdot \mathrm{d}\,\vec{S} \tag{7.14}$$

where the surface integral extends over the surface (any surface) bounded by the closed curve of the line integral, the positive direction of $\mathrm{d}\,\vec{S}$ being determined in relation to the circulation path by the right-hand rule.

The vector curl $\vec{B}$ is then a function of $\vec{B}$ at every point in space which is mathematically related to the circulation of $\vec{B}$ by this formula.

The mathematical relation between the line integral of a vector $\vec{B}$ about a closed path and the surface integral of curl $\vec{B}$ was derived by George Gabriel Stokes, and is known as Stokes' theorem. There is no physics in it. But we have seen that if $\vec{B}$ represents the magnetic field,

$$\oint \vec{B} \cdot \mathrm{d}\vec{s} = \mu_0 \int \vec{j} \cdot \mathrm{d}\vec{S} \tag{7.15}$$

where $\vec{j}$ is the current density. Therefore, for any surface $\int$ cur $\vec{B} \cdot \mathrm{d}\vec{S}$ must equal $\int \mu\vec{j} \cdot \mathrm{d}\vec{S}$, a demand which is impossible to satisfy unless

$$\operatorname{curl} \vec{B} = \mu_0 \vec{j} \tag{7.16}$$

at every point. This is the physical relation we have sought between $\vec{B}$ and the current at any point in space.

The name *curl* suggests going around, and we have arrived at the idea of curl by considering circulation, but of course curl is not identical with circulation. Consider, as a simple example, a long conducting circular cylinder in which there is a uniform current density $\vec{j}$, as indicated in the cross-section diagram (Figure 3.6). Curl $\vec{B} = \mu_0 \vec{j}$ has non-vanishing value only within the cylinder, but $\oint \vec{B} \cdot \mathrm{d}\vec{s} = I$, the total current threading the circulation loop, even if every point on the loop is outside the cylinder. There is, of course, a magnetic field $\vec{B}$ both outside and inside the cylinder, and it is $\vec{B}$ itself of which one takes the circulation.

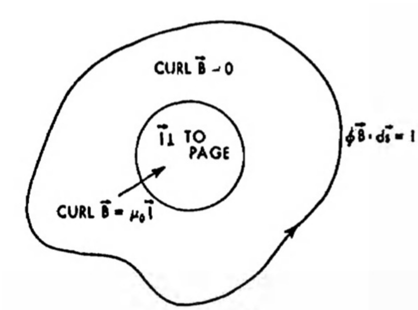


Figure 7.6: A circular cylinder with uniform current density.

The magnetic field intensity itself can be mapped out by a compass needle, a single small magnetic dipole free to orient itself along the field lines; the curl of the field intensity can be demonstrated with a magnetic "quadrupole" – two small permanent magnets with like poles cemented together – but an "octopole" is more stable and convenient. The negative (south-seeking) poles of four small permanent magnets can be cemented to a wire as indicated in Figure ??, so that their positive or N poles are the four tips of a cross at right angles to the wire. A cork attached to the wire will make the whole contrivance float in a solution of *NaCl*, for example. If the electrodes are arranged so that the current flows vertically in a cylinder of electrolyte, the "curl-$\vec{H}$-meter" will rotate continuously, the sense of rotation depending on the direction of the current. Outside the cylinder of current a dipole will show the presence of a magnetic field, but the "curl-$\vec{H}$-meter" does not rotate in the absence of current in the solution.

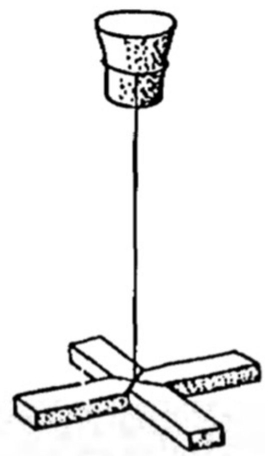


Figure 7.7: An "octopole" with four magnets.

We have noted in electrostatics that the line integral of the electric field intensity $\vec{E}$ between any two points is independent of the path connecting the points and thus $\oint \vec{E} \cdot \mathrm{d}\vec{s} = 0$ around any closed curve. It is now seen from the definition of the curl of a vector that this absence of circulation corresponds to the statement that $\operatorname{curl} \vec{E} = 0$ at every point in an electrostatic field. We also found that the point by point relation satisfied by $\vec{E}$ so as to express the physical content of Coulomb's law is $\operatorname{div} \vec{E} = \rho/\epsilon_0$, where ρ is the electric charge density. From the definition of the divergence of a vector and the fact that the net flux of the magnetic field $\vec{E}$ from any closed volume is zero, it follows that $\operatorname{div} \vec{B} = 0$. All these relations can be summarized:

ELECTROSTATICS	MAGNETOSTATICS
$\operatorname{div} \vec{E} = \dfrac{\rho}{\epsilon_0}$	$\operatorname{div} \vec{B} = 0$
$\operatorname{curl} \vec{E} = 0$	$\operatorname{curl} \vec{B} = \mu_0 \vec{j}$

These are the basic equations of electrostatics and magnetostatics. Their physical content is Coulomb's law and Ampére's law.

In electrostatics we went further, and found that the determination of the field $\vec{E}$ corresponding to any configuration of static charges was much facilitated by the introduction of a scalar potential function ϕ. The properties of the static field included the condition that the line integral of $\vec{E}$ from one point to another is given by

$$-\int_1^2 \vec{E} \cdot \mathrm{d}\vec{s} = \phi_2 - \phi_1 \tag{7.17}$$

the difference of potential between the two points, The possibility of relying on ϕ to obtain $\vec{E}$ depends on the condition that $\operatorname{curl} \vec{E} = 0$. It is left to the problems to show that the curl of the gradient of any scalar function of position vanishes identically.

It is clear that we cannot depend on a scalar potential to obtain $\vec{B}$ if the magnetic field owes its existence to currents, since the circulation of $\vec{B}$ does not in general vanish. But the divergence of $\vec{B}$ does vanish – the lines of $\vec{B}$ never begin or end. These conditions suggest that the magnetic field may be written as the curl of another vector, for it can be shown that the divergence of any vector which is itself a curl is identically zero. To show this let us again consider a finite volume (Figure 7.8), on the surface on which there is a closed

curve that divides the surface into S_1 and S_2. For any vector field $\vec{A}$ the circulation of $\vec{A}$ about the closed curve is

$$\begin{aligned}\oint \vec{A}\cdot \mathrm{d}\vec{s} &= \int \operatorname{curl}\vec{A}\cdot \mathrm{d}\vec{S}_1 \\ &= -\int \operatorname{curl}\vec{A}\cdot \mathrm{d}\vec{S}_2\end{aligned}$$

by the right-hand rule, since we have taken $\mathrm{d}\vec{S}$ positive outward from the enclosed volume for both surfaces.

But the total flux of curl $\vec{A}$ out of the volume is

$$\begin{aligned}\int \operatorname{curl}\vec{A}\cdot \mathrm{d}\vec{S}_1 + \int \operatorname{curl}\vec{A}\cdot \mathrm{d}\vec{S}_2 &= \int_{\text{S closed}} \operatorname{curl}\vec{A}\cdot \mathrm{d}\vec{S} \\ &= \int \vec{A}\cdot \mathrm{d}\vec{s} - \int \vec{A}\cdot \mathrm{d}\vec{s} \\ &= 0\end{aligned}$$

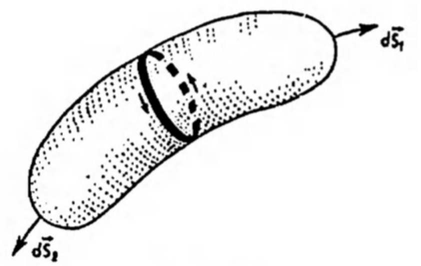


Figure 7.8: Circulation in a finite volume.

This is true for any volume, and for any closed curve on the surface of the volume, and must therefore be true in the limit:

$$\operatorname{div}\operatorname{curl}\vec{A} = \lim_{\Delta V\to 0} \frac{\int_{\text{S closed}} \operatorname{curl}\vec{A}\cdot \mathrm{d}\vec{S}}{\Delta V} \equiv 0$$

Thus writing $\vec{B}$ as the curl of another vector ensures that its divergence is zero, a condition imposed physically on the magnetic field $\vec{B}$ if it is produced by currents, and defined as a property of $\vec{B}$ as produced by magnets to include the inseparability of magnetic poles along with Coulomb's law for magnets.

In view of the limiting process by which the curl was defined it is not surprising that it resembles the gradient and the divergence in being a differential operator with respect to coordinates in three-dimensional space. In order to make quantitative use of the concept we must write it in terms of coordinates, although the physical quantity it represents is quite independent of the particular coordinate system chosen. For our purposes, the familiar Cartesian coordinates will suffice, particularly if we remember to choose the origin of coordinates and the orientation of the axes so as to make the description of the physical problem as simple as possible.

Let us consider the x component of curl $\vec{A}$ in a right-handed Cartesian coordinate system at the point (x, y, z). In Figure 7.9, $\mathrm{d}y$ and $\mathrm{d}z$ are shown as finite increments in the direction of increasing

y and z; eventually we shall let $\mathrm{d}\,y$ and $\mathrm{d}\,z$ become as small as we please. By definition,

$$(\operatorname{curl}\vec{A})_x = \lim_{\mathrm{d}\,z\,\mathrm{d}\,y \to 0} \frac{1}{\mathrm{d}\,z\,\mathrm{d}\,y} \oint \vec{A} \cdot \mathrm{d}\,\vec{S}_1 \tag{7.18}$$

with the line integral taken around the boundary of the small rectangle shown.

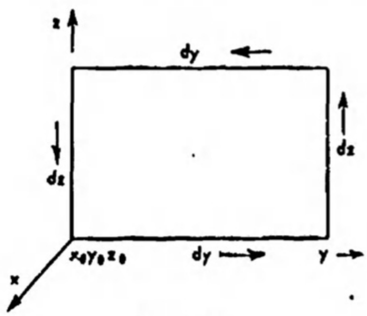


Figure 7.9: Curl in a right handed coordinate system.

The vector $\vec{A}\,(x,\, y,\, z)$ must vary with changing y or z (or both) if the line integral is to be different from zero, and we must allow for this variation to first order in $\mathrm{d}\,y$ and $\mathrm{d}\,z$. For the legs of the rectangle adjacent to the point $x_0,\, y_0,\, z_0$. $\vec{A}$ is $\vec{A}(x_0,\, y_0, z_0)$, but all of the leg $\mathrm{d}\,z$ on the right is at $y_0 + \mathrm{d}\,y$ and the leg $\mathrm{d}\,y$ at the top is at $z_0 + \mathrm{d}\,z$. The line integral is then

$$\begin{aligned} A_y\,(x_0,\, y_0,\, z_0)\,\mathrm{d}y + A_z\,(x_0,\, y_0 + \mathrm{d}\,y,\, z_0)\,\mathrm{d}z \\ - A_y\,(x_0,\, y_0,\, z_0 + \mathrm{d}z)\,\mathrm{d}y \\ - A_z\,(x_0,\, y_0,\, z_0)\,\mathrm{d}z \end{aligned}$$

the last two terms being negative because the path is traversed in the direction of decreasing y and z, respectively. To take into account the fact that in one term A_z is evaluated at $y_0 + \mathrm{d}y$, we write

$$A_z\,(x_0,\, y_0 + \mathrm{d}y,\, z_0) = A_z\,(x_0,\, y_0,\, z_0) + \left(\frac{\partial A_z\,(x_0,\, y_0,\, z_0)}{\partial y}\right)_{y_0} \mathrm{d}y$$

where $(\partial A/\partial y)_{y_0})$ is the slope of A_z plotted against y evaluated at the point y_0, the other two coordinates remaining unchanged. This does not imply that A_z is a linear function of y. If $\mathrm{d}y$ were to remain a finite length we should have to worry in more detail about the dependence of A_z on y. Similarly,

$$A_y\,(x_0,\, y_0,\, z_0 + \mathrm{d}z) = A_y\,(x_0,\, y_0,\, z_0) + \left(\frac{\partial A_y\,(x_0,\, y_0,\, z_0)}{\partial z}\right)_{z_0} \mathrm{d}z$$

When these expressions are substituted in the closed line integal all terms which do not involve derivatives cancel, and we are left with

$$(\operatorname{curl}\vec{A})_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \tag{7.19}$$

where the coordinates need not be written explicitly.

The other components of $\operatorname{curl}\vec{A}$ can be derived in the same way, but it is equally valid to invoke the symmetry of a right-handed Cartesian coordinate system and obtain the y and z components by cyclic permutation of x, y, z:

$$\begin{aligned}(\operatorname{curl}\vec{A})_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\\ (\operatorname{curl}\vec{A})_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\end{aligned} \tag{7.20}$$

The result is reminiscent of the form of the cross product of two vectors. If $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit vectors in the direction of increasing x, y, z, we recall that

$$\vec{A} \times \vec{B} = (A_y B_z - A_z B_y)\,\hat{x} + (A_z B_x - A_x B_z)\,\hat{y} + (A_x B_y - A_y B_x)\,\hat{z}$$

In writing the gradient and the divergence in Cartesian coordinates we have already made use of a vector differential operator

$$\boldsymbol{\nabla}\cdot \equiv \left(\hat{x}\frac{\partial}{\partial x}\right) + \left(\hat{y}\frac{\partial}{\partial y}\right) + \left(\hat{z}\frac{\partial}{\partial z}\right)$$

and have found it convenient to write $\operatorname{grad}\phi = \boldsymbol{\nabla}\phi$, and $\operatorname{div}\vec{E} = \boldsymbol{\nabla}\cdot\vec{E}$. Here we may write $\operatorname{curl}\vec{A} = \boldsymbol{\nabla}\times\vec{A}$, and the determinantal form of the cross product of two vectors is again a helpful mnemonic device:

$$\operatorname{curl}\vec{A} = \boldsymbol{\nabla}\times\vec{A} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} \tag{7.21}$$

The symbol $\boldsymbol{\nabla}\cdot$ (del) is useful in the manipulation of mathematical relations between physical concepts, since it can be treated as an ordinary vector so long as its role as a differential operator is kept firmly in mind, but the meaning is more apparent if we say "curl" instead of "del cross" in reading a formula.

We have seen that the magnetic field $\vec{B}$ can be written as the curl of some other vector which is also defined for all points of space. The relation between a vector and its curl is clarified by consideration of some simple examples. Let us take

$$\vec{B} = \operatorname{curl}\vec{A} \tag{7.22}$$

where $\vec{A} = B_0(\hat{y}\,x - \hat{x}\,y)$. It is easily found that $\operatorname{curl}\vec{A} = B_0\vec{z}$, a uniform field parallel to the z axis of coordinates, but what about

the lines of $\vec{A}$? It is a simple exercise to show that they are concentric circles, lying in planes perpendicular to the z axis.

As a second example consider $\vec{A} = \mu_0 j/4\pi(x^2 + y^2)\hat{z}$ parallel to the z axis but depending symmetrically on x and y. (The first factor is constant and a scalar.) Find $\vec{B} = \text{curl}\vec{A}$. Now find curl$\vec{B}$. The answer has been anticipated in writing the constant factor in the expression for $\vec{A}$, but the relation between the succession of vectors found by taking the curl is interesting.

The vector $\vec{A}$ of which the magnetic field $\vec{B}$ is the curl is called the "vector potential." The scalar potential in electrostatics was defined as work per unit charge, measured in volts, and could be traced to its sources by summing the effects of all charges giving rise to the field intensity $\vec{E}$. We saw that its gradient is just the electric field intensity (except that we change the sign), that is, the physically observable force per unit charge is derived from the scalar potential by means of the operator "del." The relation between work and force is then quite apparent, and the measurement of $\vec{E}$ in volts per meter reinforces the connection. The role of the vector potential in magnetic fields is more complicated, largely because the force per unit current element is not in the direction of the field B but at right angles to it . The simplest justification for using the word potential here is that $\vec{A}$ represents a quantity whose derivative (its curl, this time) is a physically measurable field, namely, $\vec{B}$.

Just as the scalar potential ϕ can be traced to electric charges we should expect that $\vec{A}$ can be traced to currents. We shall write down the correct relation between the vector potential and linear current sources, then show that this relation is compatible with the dependence of $\vec{B}$ on these same currents, as known from Ampére's law. Before doing so, however, we should note that $\vec{A}$ is not completely determined by the demand that its curl give the correct magnetic field; any vector whose curl is zero could be added to $\vec{A}$ without affecting $\vec{B}$ at all.

We have noted in the description of the fields $\vec{E}$ and $\vec{B}$ that it is necessary to know both the curl and the divergence to specify a vector. Since $\vec{A}$ has been introduced only so that its curl represents $\vec{B}$, nothing has been said about its divergence, which may be anything. It is customary in magnetostatics to require that div$\vec{A} = 0$, but this restriction is arbitrary. The fact that the vector potential is not completely defined by requiring that its curl give the right magnetic field is reminiscent of the ambiguity of the scalar potential

of electrostatics, to which any arbitrary constant could be added.

It is possible to show that the expression for $\vec{B}$ in terms of a current, (7.2), may be written as the curl of some other vector quantity which can then be Identified as the vector potential. It is somewhat simpler, mathematically, to write down a formula for the vector potential and show that its curl gives (7.2).

Let us put

$$\vec{A} = \frac{\mu_0}{4\pi} \oint \frac{I\,\mathrm{d}\vec{s}}{r} \tag{7.23}$$

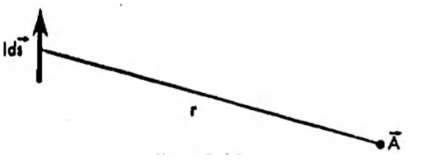


Figure 7.10: Finding curl at a point.

where $I\,\mathrm{d}\vec{s}$ is an element of current, as usual, and r is the distance from $I\,\mathrm{d}\vec{s}$ to the point where $\vec{A}$ (and hence $\vec{B}$) is to be computed, which we may call the field point (see Figure 7.10). The curl of $\vec{A}$ is to be taken *at* the field point, and depends on the coordinates of that point, not those of the source. (After all, the same field $\vec{B}$ at some point could be produced equally well by a variety of source configurations.) Moreover, owing to the principle of superposition, it makes no difference whether we take the curl at each point of the integrand and then sum, or first sum over all parts of the circuit $I\,\mathrm{d}\vec{s}$ and then take the curl at the field point. In other words,

$$\begin{aligned}
\operatorname{curl}\vec{A} &= \frac{\mu_0}{4\pi}\boldsymbol{\nabla}\times\oint\frac{I\,\mathrm{d}\vec{s}}{r}\\
&= \frac{\mu_0}{4\pi}\oint\boldsymbol{\nabla}\times\left(\frac{I\,\mathrm{d}\vec{s}}{r}\right)\\
&= \frac{\mu_0}{4\pi}\oint\boldsymbol{\nabla}\cdot\left(\frac{1}{r}\right)\times I\,\mathrm{d}\vec{s}
\end{aligned}$$

where in the last term we have made explicit use of the fact that the del operator does not act on the coordinates of the current element. The field point is involved in the integrand only through the factor $1/r$, where r is the distance from $I\,\mathrm{d}\vec{s}$ to the field point, and thus depends both on the variable of integration $\mathrm{d}\vec{s}$ and those at which the vector derivative is taken.

In writing curl $\vec{A}$ in the final form above we have taken advantage of tho fact that tho operator del behaves like a vector as well as a differential operator, but the integrand now reads differently: $\boldsymbol{\nabla}\cdot(1/r) = \operatorname{grad}(1/r)$, and we have the cross product of a gradient of a scalar and $I\,\mathrm{d}\vec{s}$. The gradient of $1/r$ is very familiar from electrostatics, since the electrostatic potential of a point charge is proportional to $1/r$, where r is the distance from the point charge to the point at which we take the gradient to find $\vec{E}$, the electric field intensity. (Of course, we could also simply compute it again.)

If we take the source point, the position of $I\,\mathrm{d}\vec{s}$, as the origin of coordinates, grad $(1/r) = -\hat{r}/r^2$, where $\hat{r}$ is a unit vector directed from $I\,\mathrm{d}\vec{s}$ to the field point. With this substitution, and a change of order in the cross product which changes its sign,

$$\mathrm{curl}\,\vec{A} = \frac{\mu_0}{4\pi}\oint \frac{I\,\mathrm{d}\vec{s}\times\hat{r}}{r^2} \tag{7.24}$$

which is identical with (3.2). Thus our expression for A is justified.

The vector potential, like the scalar potential, is defined at the field point, but we see that it is very closely related to the current. In fact, for each element of current

$$\Delta\,\vec{A} = \frac{\mu_0}{4\pi}\frac{I\,\mathrm{d}\vec{s}}{r} \tag{7.25}$$

and thus each increment of $\vec{A}$ is in the same direction as the current element which produces it. From the definition of the curl we see again that the field $\vec{B}$ is at right angles to $\vec{A}$, as well as to $I\,\mathrm{d}\vec{s}$.

The vector potential is sometimes useful in solving problems in magnetostatics, but it does not play nearly so practical a role in determining $\vec{B}$ from steady currents as does the scalar potential in electrostatics problems. On the other hand, it is almost indispensible in relating fields to non-steady currents, the fields produced by time varying currents. We shall return to this point in another chapter. But before leaving the subject, let us note an interesting relation between the vector potential and the flux of the magnetic field $\vec{B}$ through a surface.

Thus far we have considered the circulation of $\vec{B}$ in relation to the current through a surface bounded by the circulation path, and have noted that the flux of $\vec{B}$ through a *closed* surface always vanishes, but we can now derive a new circulation law. Consider the flux of $\vec{B}$ through a surface bounded by a closed path. By definition this flux is the integral of $\vec{B}\cdot\mathrm{d}\vec{S}$ over the surface. But $\vec{B} = \mathrm{curl}\,\vec{A}$. Therefore the flux through the surface is

$$\phi_B = \int \vec{B}\cdot\mathrm{d}\vec{S} = \int \nabla\times\vec{A}\cdot\mathrm{d}\vec{S} = \oint \vec{A}\cdot\mathrm{d}\vec{S} \tag{7.26}$$

just the line integral of $\vec{A}$ round the boundary of the surface. In the early chapters of Monograph III you will have learned that a *changing* flux of $\vec{B}$ through a surface is accompanied by a circulation

of the electric field intensity $\vec{E}$. Thus

$$\frac{\mathrm{d}\phi_B}{\mathrm{d}t} = \int \frac{\partial \vec{B}}{\partial t} \cdot \mathrm{d}\vec{S} = -\oint \vec{E} \cdot \mathrm{d}\vec{s} = \oint \frac{\partial \vec{A}}{\partial t} \cdot \mathrm{d}\vec{s} \tag{7.27}$$

and therefore an electric field which has a circulation is related to the vector potential:

$$\vec{E} = -\left(\frac{\partial \vec{A}}{\partial t}\right)$$

The reformulation of magnetostatics in terms of vector calculus has in itself added nothing to tho physical content of Ampére's law; in fact, the physics of curl $\vec{H} = \vec{j}$ may be less transparent than $\oint \vec{H} \cdot \mathrm{d}\vec{s} = I$, except for the "curl-$\vec{H}$-meter" which works only in fluid conductors. As for the vector potential, $\vec{A}$ is sometimes, but not always, useful for solving problems, but it is not even directly observable by means of classical currents or magnets. Tho power of the differential formulation of the laws of electricity and magnetism is fully realized only when variations of the fields in time are taken into account. If $\vec{E}$ and $\vec{B}$ (or $\vec{H}$) are permitted to vary in time, as they are bound to do, a whole new set of electromagnetic consequences are observed, Tho differential forms of Ampére's and Faraday's laws helped Maxwell to conclude that "light itself (including radiant heat, and other radiations if any) is an electromagnetic (disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws." Even Maxwell did not succeed in tracing electromagnetic radiation to its sources; to accomplish this in an unambiguous way requires the vector potential, or something equivalent to it, a single quantity to which both the electric field and the magnetic field are related.

Part III

Circulation Laws and Their Consequences

Sections 8.4 through 8.7 can be omitted without losing the line of the argument leading to the propagation of electromagnetic waves. Nothing in Chapters 9 and 10 depends explicitly on the development in these sections.

I had hoped to complete two additional sections: Experimental Confirmation and Reference Frames. These would have appeared at the end of Chapter 10. The first would have described the experiments confirming Maxwell's predictions about the propagation of electromagnetic waves. The second would have described the interplay of $q\vec{E}$ and $qv \times \vec{B}$ forces as we transform from one inertial frame to another, and it would have exhibited the difficulties inherent in applying Galilean relativity to Maxwell's equations. While I had no intention of introducing special relativity in any detail, this seemed the ideal place at which to set the stage. In any case, students should not be left with the impression that everything they learned about transformations in the context of Newtonian mechanics can be carried over to electromagnetic theory.

R.T. Mara

Chapter 8

Faraday's Law Of Induction

8.1 Faraday's Discovery

Michael Faraday knew about Ampére's work, so he knew that a magnetic field accompanies a current. The monograph Magnetostatics gives a thorough discussion of the relationship between magnetic fields and currents. In mathematical terms,

$$\vec{B} = \sum \Delta\vec{B} = \frac{\mu_0}{4\pi} \sum_{s} \frac{I\,\Delta\vec{s} \times \hat{r}}{r^2} \tag{8.1}$$

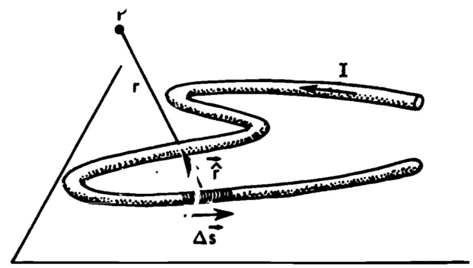


Figure 8.1: The quantities appearing in equation (8.1) describing the vector $\vec{B}$ at the field point P. Only a part of the complete circuit is shown.

where $\vec{B}$ is the magnetic induction field at a point P (see Figure 8.1), I is the steady current in the element of length $\Delta\vec{s}$ along the circuit and $\Delta\vec{s}$ is in the direction of that current, r is the distance between $\Delta\vec{s}$ and the field point P, and $\hat{r}$ is a unit vector pointing from $\Delta\vec{s}$ toward P.

There is a circulation law closely associated with equation (8.1).

$$\sum_{\text{s closed}} \vec{B}\,\Delta\vec{s} = \mu_0 I = \mu_0 \sum_{s} \vec{j} \cdot \Delta\vec{S} \tag{8.2}$$

This says that the circulation of $\vec{B}$ is proportional to the net charge that passes per second through any surface S bounded by the circulation path s. There is a convention that relates the sense in which $\Delta\vec{S}$ is taken as positive around the circulation path and the direction in which $\Delta\vec{S}$ and hence I is to be considered positive. This convention is illustrated in Figure 8.2.

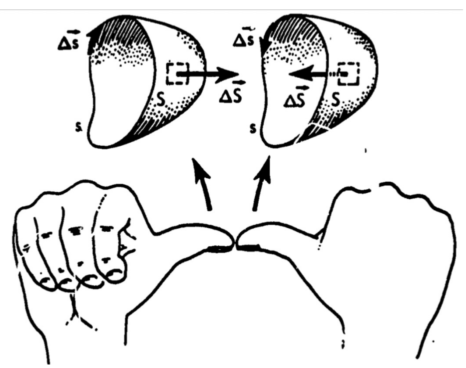


Figure 8.2: The right-handed convention relating the sense of $\Delta\vec{s}$ and the direction of $\Delta\vec{S}$ for Ampére's circulation law.

The monograph Magnetostatics considers $\vec{B}$ only in cases for which the field point is in empty space, outside of matter, and equations (8.1) and (8.2) are valid only for such cases. We shall also study fields in empty space in this monograph.

While Faraday did not have the mathematical conception of Ampere's work that appears in equations (8.1) and (8.2), he understood the physics contained in those equations. And he had a strong hunch that there was more to the story. To understand Faraday's language and hence something of the way he thought, we quote the first paragraph of a paper he presented to the Royal Society in 1831.

1. The power which electricity of tension (electrostatic charge) possesses of causing an opposite electrical state in its vicinity has been expressed by the general term induction; which, as it has been received into scientific language, may also, with propriety, be used in the same general sense to express the power which electrical currents may possess of inducing any particular state upon matter in their immediate neighborhood, otherwise indifferent. It if with this meaning that I purpose using it in the present paper.[1]

In much the same sense that he imagines static charge *induces* opposite charges in nearby conductors, Faraday sees a current inducing a magnetic field in its neighborhood. Faraday continues in the same paper.

2. Certain effects of the induction of electrical currents have already been recognized and described : as those of magnetization; … Still it appeared unlikely that these could be all the effects which induction by currents could produce; especially as, upon dispensing with iron, almost the whole of them disappear, whilst yet an infinity of bodies, exhibiting definite phenomena of induction with electricity of tension, still remain to be acted upon by the induction of electricity in motion.
3. Further: Whether Ampére's beautiful theory were adopted, or any other, or whatever reservation were mentally made, still it appeared very extraordinary, that as every electric current was accompanied by a corresponding intensity of magnetic action at right angles to the current, good conductors of electricity, when placed within the sphere of this action, should not have any current induced through them, or some sensible effect produced equivalent in force to such current.
4. These considerations, with their consequences, the hope of obtaining electricity from ordinary magnetism, have stimulated me at various times to investigate experimentally the inductive effect of electric currents

Faraday apparently felt deeply that a current in one conductor

[1] Quoted from *Faraday's Experimental Researches In Electricity, Volume I*, which is the first of three volumes in which are collected all the papers Faraday published in the *Philosophical Transactions* in the years 1831–1838 (Richard and John Edward Taylor, London, 1839).

should do something to a nearby conductor. And paragraph 4 above seems to indicate Faraday's hunch that if magnetic could result from a current, then, in some way or other, current should be obtainable from magnetism. His viewpoint is at least partly evident in experiments he performed as early as 1824. He passed a magnet through a conducting helix but noted no effect from doing it. He passed a current through one wire but found nothing in a nearby wire. Then, in ten days of experimentation, starting August 29, 1831, and ending November 4, 1831, Faraday found what he was after, although not really what he had expected. On November 24, he read his famous paper to the Royal Society, and the quotations above are taken from that paper.

In that short san Faraday discovered just about every way possible to induce current in conductors, and he formulated a law that accounted for all of them. That work is the basis for all modern electric power, from the giant turbines, at power stations to the motors that drive lathes and drill presses. We do not intend to cover all the work Faraday reported in even that first paper. In any case, no one is likely to improve on Faraday's own report, so James Clerk Maxwell's advice is good,

> I would recommend the student, after he has learned, experimentally if possible, what are the phenomena to be observed, to read carefully Faraday's *Experimental Researches in Electricity*. He will there find a strictly contemporary historical account of some of the greatest electrical discoveries and investigations, carried on in an order and succession which could hardly have been improved if the results had been known from the first, and expressed in the language of a man who devoted much of his attention to the methods of accurately describing scientific operations and their results.[2]

In his laboratory notebook Faraday recorded what is apparently the first experiment exhibiting the effect that now carries his name.

> I have had an iron ring made (soft iron), iron round and 7/8$^{\text{ths}}$ of an inch thick, and ring six inch inches in external diameter. Wound many coils of copper round, one half of the coils being separated by twine and calico; there were three lengths of wire, each about twenty four feet long, and they could be connected as one length or used as separate lengths. By trials with a trough each was insulated from the other. Will call this side of ring *A*. On the other side, but separated by an interval , was wound wire in two pieces, together amounting to about sixty feet in length, the direction being as with the former coils. This side call *B*.
>
> Charged a battery of ten pairs of plates four inches square. Made the coil on *B* side one coil, and connected its extremit-

[2] Quoted from the Preface to the First Edition of James Clark Maxwell's *A Treatise on Electricity and Magnetism, Volume I*, 3rd edition, 1891, republished by Dover Publications, Inc., New York (1934).

> ies by a copper passing to a distance, and just over a magnetic needle (three feet from wire ring), then connected the ends of one of the pieces on *A* side with battery: immediately a sensible effect on needle. It oscillated and settled at last in original position. On breaking connection on *A* side of battery, again a disturbance of the needle.[3]

[3] Quoted from *The Life and Letters of Faraday, Volume II*, 2nd edition, revised by Bence Jones; Longmans, Green, and Co. (1870).

Figure 8.3 shows Faraday's setup. It is not difficult to depict, given the clear account in the notebook. The important thing Faraday describes is that the magnetic needle is disturbed only at the instant the contact to the battery is *made* or *broken*; in between it settles down, even though there is current in *A*.

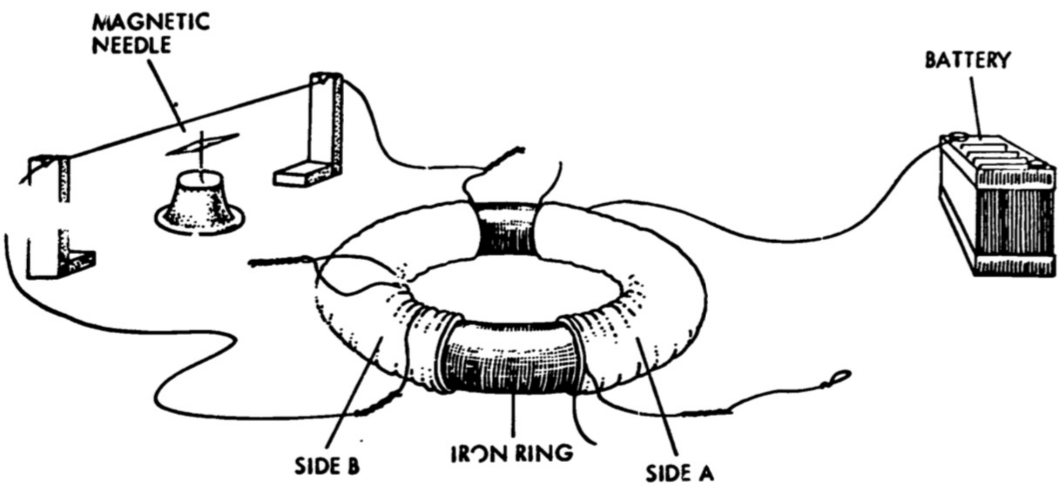


Figure 8.3: The experimental setup when Faraday first noticed an induction effect.

While this was likely the first experiment in which Faraday got a noticeable effect, it is not the first he describes in his paper. There he describes his results, "not as they were obtained, but in such a manner as to give the most concise view of the whole." We shall select a few of his experiments to explore in this chapter, but we shall not return to the one already described. For our purposes here, the iron ring is just a complication. When you have learned something about fields in magnetic materials, you will understand why the effect is considerably stronger with the iron present, and then you can make reasonable conjectures about why Faraday first noticed it this way. The first experiment Faraday described in his paper is equivalent to the one pictured in Figure 8.3, but there is no iron present. Figure 8.4 (see next page) pictures the way Faraday's worktable might have been arranged when he performed this experiment.

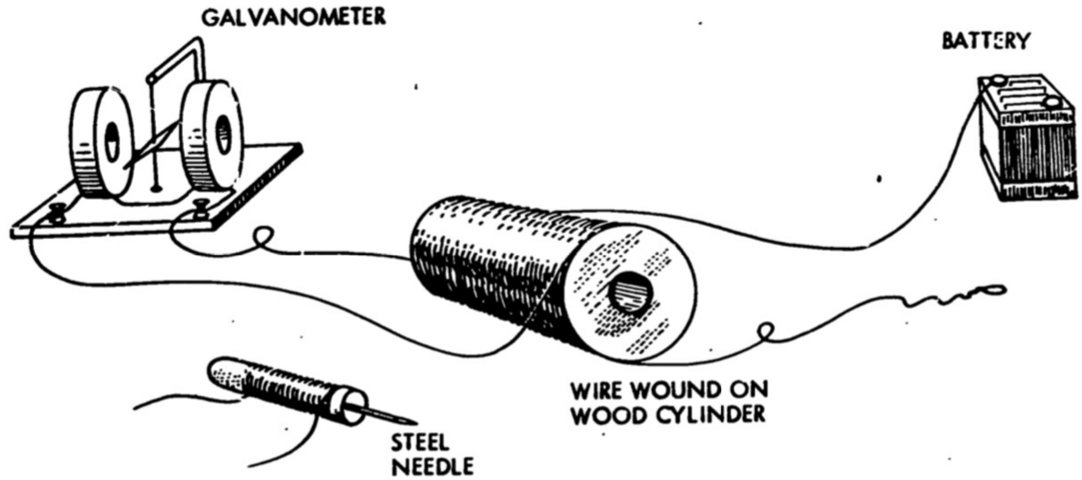


Figure 8.4: Faraday's arrangement for detecting induced currents.

He wound copper wire around a wooden cylinder, and i.e. wound thread at the same time to keep successive turns of wire separated. He covered this layer with cloth, and wound a second layer of wire and thread on top of the first. He continued until he had a total of twelve windings, each wound in the sane sense . He then connected the ends of the first, third, fifth, etc., windings to make a single conductor. He did the same with the second, fourth, sixth, etc. In this way "two principal helices were produced, closely interposed, having the same direction, not touching anywhere, and each containing one hundred and fifty-five feet length of wire."

He connected one principal helix across a galvanometer and the other across "a voltaic battery of ten pairs of plates four inches square, with double coppers and well charged; yet not the slightest sensible deflection of tho galvanometer needle could be observed." He did the same thing with other metals used for the wire. No difference. He then made a bigger version of the whole affair. He increased the length of wire in each principal helix to two hundred feet and the battery to "one hundred pairs of plates four inches square, with double coppers, and well charged." This time something did happen.

> When the contact was made, there was a sudden and very slight effect at the galvanometer, and there was also a similar slight effect when the contact with the battery was broken. But whilst the voltaic current was continuing to pass through the one helix, no galvanometer appearances nor any effect like induction upon the other helix could be perceived, ...

Faraday was certain that the same effect must have been present

when he had used the smaller coils and weaker battery. He suspected that his galvanometer was simply not sensitive enough to detect it, so he devised an ingenious substitute for the galvanometer. He wound a coil around a glass tube and placed a steel needle inside the tube (see Figure 8.4). If a current passes through that coil, then there will be a magnetic field inside the tube that will magnetize the needle. He substituted this arrangement for the galvanometer and repeated the experiment.

With the needle originally unmagnetized, he made the connection to the battery in the *primary* circuit. Before breaking that connection he pulled out the needle and found it to be magnetized. Now he put a second, unmagnetized needle into the tube and then broke the primary circuit. The second needle was also magnetized, but its polarity was the reverse of that found in the first needle.

When he put in an unmagnetized needle before closing the primary circuit and left it in until after that circuit was broken, he found "little or no magnetism." When he closed the primary circuit before putting the needle in the tube, and then removed the needle before breaking that circuit, he found no magnetism in the needle.

These results can be explained only on the basis that the current in the induced, or secondary, circuit is in one direction for a short time when the primary has just been closed and in the opposite direction for a short time when the primary has just been broken, and that there is no current in the secondary during the intervening time. That is, current appears in the secondary only when the current in the primary is *changing*.

Further, Faraday discovered that "The [galvanometer] deflection on making a battery contact always indicated an induced current in the opposite direction to that from the battery; but on breaking the contact the deflection indicated an induced current in the same direction as that of the battery."

In Figure 8.5 we have a summary of the experimental results. The experimental setup is a simplified version of Faraday's, so that the effects can be illustrated more clearly.

What is it that is changing at the secondary that brings about the induced current? Well, as the current builds up in the primary, a *magnetic field* is building up, too. Can it be that a changing magnetic field causes the induced current? There is a simple way to find out.

The experimental setup shown in Figure 8.5 can be modified as shown in Figure 8.6. A movable magnet substitutes for the primary circuit. When the magnet moves, the magnetic field at the secondary

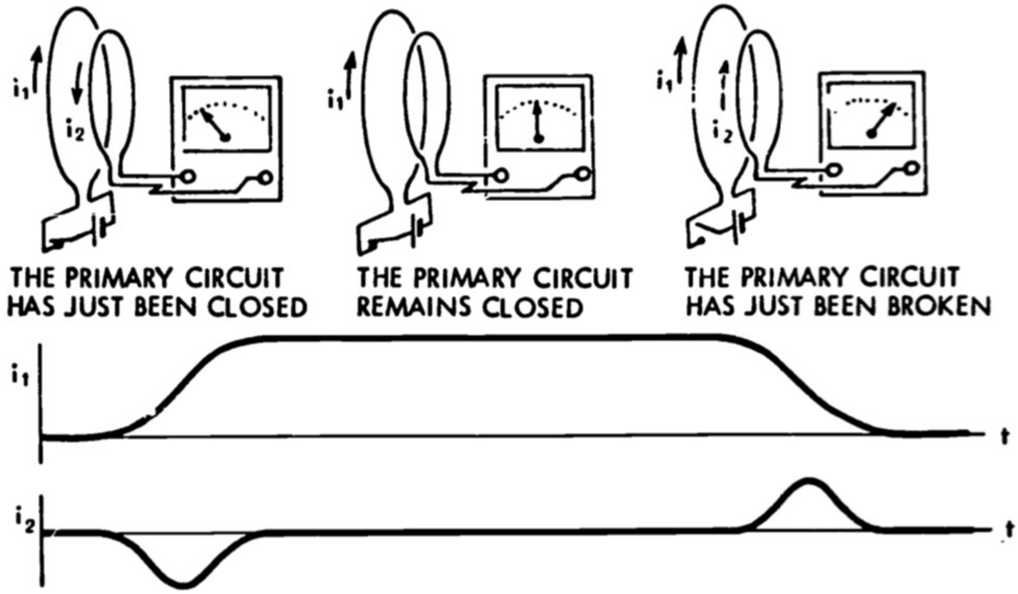


Figure 8.5: Changing the current in a nearby circuit can induce a current.

certainly changes. If a changing magnetic field is responsible for induced currents, a current should appear in the secondary while the magnet is moving. And it does, as Faraday discovered in just about this way.

Now we have an interesting question. Suppose that in the experiment depicted in Figure 8.6 we arrange to have the conducting loop move and the magnet remain at rest (see Figure 8.7). If we can extend what we have learned in mechanics into the realm of electricity and magnetism, then the physics should be the same, no matter which moves.

And that turns out to be correct. The galvanometer response is the same whether the magnet moves to the left and the loop is stationary or the magnet is stationary and the loop moves to the right. And Faraday discovered that, too.

At this point, Faraday had his hands on what he wanted. Things had not turned out to be quite what he had imagined when he began his *Researches*, but his hunch was correct. His hope that a "current should be obtainable from magnetism" was realized.

Note: The experiments described in this section are relatively easy to do, and reading about them is not really the same as doing them or, at least, seeing them done. Maxwell in the Preface to his *Treatise on Electricity and Magnetism*, says, "… before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's *Experimental Researches in*

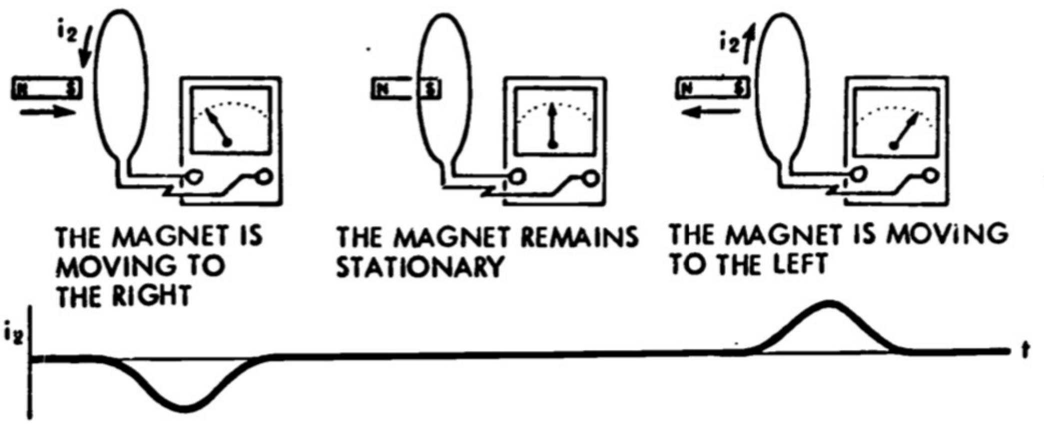


Figure 8.6: Moving a nearby magnet can induce a current.

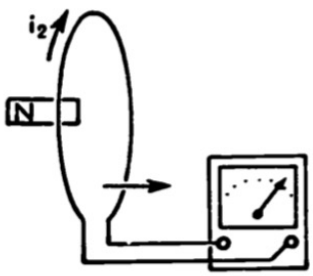


Figure 8.7: Moving the loop in a fixed magnetic field can induce a current.

Electricity ... I would recommend the student, after he has learned, experimentally if possible, what are the phenomena to be observed, to read carefully Faraday's *Experimental Researches in Electricity*."

This chapter is certainly no satisfactory substitute for the *Researches*, and Maxwell's admonition about experimental understanding is even more germane if a student is relying largely on what is written here. It is possible to reproduce much of what Faraday did with inexpensive equipment, and it is well worth while doing just that.

8.2 Faraday's Law Of Induction

Before we proceed to the general statement of Faraday's law of induction, we want to give some simplified order to the results of Faraday's experiments. In essence, he found three ways to produce induced currents, and these are represented in Figure 8.8. He produced an induced current in a loop by changing the current in a nearby circuit. An idealized version of that experiment is depicted in Figure 8.8 (a), where the current is being changed by changing the resistance in the rheostat R. He also found that an induced current can appear in a loop when the loop is moved in a magnetic field. That experiment is depicted in Figure 8.8 (b), where a circuit with constant current is used as the source of the magnetic field. And finally, Faraday found that an induced current can appear in a stationary loop when the source of the magnetic field is moved (see Figure 8.8 (c)).

We intend to discuss all three of these experiments of Faraday's

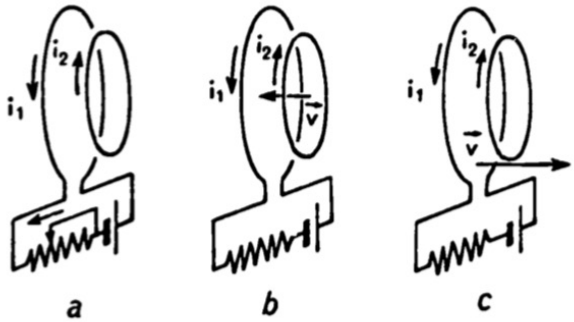


Figure 8.8: Three ways to produce an induced current.

in this chapter, but we start with a special case of the first. Suppose we put a loop inside a long coil in which we can vary the current (see Figure 8.9). If the current in the coil rises from zero to some steady value I_0 in the time Δt, then the galvanometer needle will be deflected during that time. Suppose we make up a set of loops identical in size, but constructed of different known non-magnetic conductors.

Now we use one loop at a time, and note the galvanometer deflection for each as the current in the coil rises to I_0. In general, the deflections are not the same, the deflection being greatest for those conductors whose resistance is least. This means that while the geometric arrangement is the same in each case, and while the current rises in the coil in the same way in each case, the induced current in the various loops is not the same.

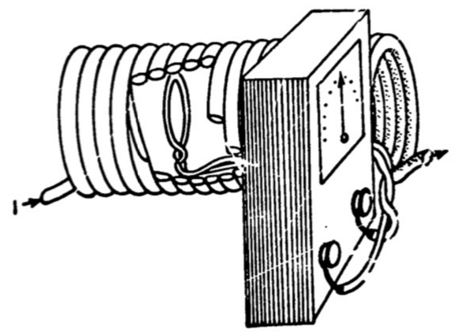


Figure 8.9: Experimental arrangement for studying induction.

We make now a different set of loops, these to have different radii but the same resistance. Again the galvanometer deflections are different, the loop with the larger radius giving the greater deflection. And finally, for a given loop, the more rapidly we increase the current in the coil the larger induced current we find in the loop. (If you try to perform this set of experiments, you will likely do better if you substitute an oscilloscope for the galvanometer.)

If we can use Ohm's law to describe what is going on in the loop, we see that while there is an induced current in the loop there must also be an induced *electromotive force*. That is,

$$\mathcal{E} = iR \tag{8.3}$$

where $\mathcal{E}$ is the electromotive force, I the current, and R the resistance of the loop.[4] Recall that the electromotive force (almost always abbreviated as *emf*) is defined as being the work done on a unit charge as that charge travels once around the circuit. It too bad that this is called a force when it really is work, but historical

[4] It is certainly true that Ohm's law is really valid only for steady currents, and the induced currents are not steady here. That means that what consequences of Ohm's law we use will in detail be incorrect, because we have not taken account of the magnetic field created by the induced current itself. Those interested in this matter can, after finishing this section, look up the meaning of the word *self-inductance*.

origins of words often color our language in strange ways

Up to this point, we have been concentrating on the detection and measurement of induced currents in describing experiments. But it turns out that the relationship between what goes on in the loop and what is happening to the environment in which that loop sits is most directly described in terms of the induced *emf*. The mathematical statement of that relationship is simply

$$\mathcal{E} = -\frac{\Delta\Phi_B}{\Delta t} \tag{8.4}$$

where $\mathcal{E}$ is the induced *emf* in the loop, and Φ_B is the flux of $\vec{B}$ threading that loop, i.e.,

$$\Phi_B = \sum_S \vec{B} \cdot \Delta\vec{S} \tag{8.5}$$

where S is any surface bounded by the loop. In words, this says that the induced *emf* $\mathcal{E}$ in a loop is equal to the negative of the time rate of change of the magnetic flux threading the loop.

While Maxwell was the first to put this simple but powerful concept into mathematical form, Faraday was the first to discover the effect. It is most often called Faraday's *law of induction*. Later on we shall call it his *circulation law*; that it is in fact a circulation law will soon be clear.

We should look closely at Faraday's law as given by equations (8.4) and (8.5), so that some important details become clear. First, we look at the reason for the minus sign in equation (8.4). We want the convention relating the sense of the *emf* and the direction of the flux change to be the same as that used in Ampére's law, where the sense in which a circulation about a closed path is related to the direction in which the current threading that path is considered to be positive (see equation (8.2) and Figure 8.2).

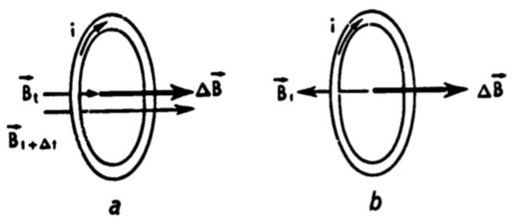


Figure 8.10: Illustration of Lenz's law. (a) $\vec{B}_t$ is the magnetic induction field at the time t, $B_{t+\Delta t}$ at the time $t+\Delta t$, and i the induced current, (b) $\vec{B}_i$ is the magnetic induction field created by the induced current.

In Figure 8.10 (a), we show a stationary loop through which Φ_B is changing. The direction of that change is the direction of $\Delta\vec{B}$.

The corresponding sense of the induced *emf* and of the induced current appear, too, as they are experimentally determined. But the relationship between these is just the opposite of our convention, so we need the minus sign in (8.4) to make that fact explicit.

There is another way to remember the sense of the induced current and, hence, of the induced *emf*. This way has a sound physical basis. The current induced in the loop is always in the sense such that the magnetic induction $\vec{B}_i$ that it creates is in the direction

that opposes the change Φ_B threading the loop. Figure 8.10 (b) illustrates this relationship, which is often called *Lenz's law*. We can see that the world must behave that way when we consider what would happen if the induced current went the other way.

If the induced current were such as to create a $\vec{B}_i$ that increases the inducing $\Delta\Phi_B$, then that increase would bring about a yet greater induced current, which would further increase $\Delta\Phi_B$, which would again increase the induced current, and so on without end. That certainly cannot happen, so the induced current must create a magnetic field that reduces the flux change $\Delta\Phi_B$. That is, Lenz's law is a special consequence of the requirement that energy be conserved. And so we understand the appearance of the minus sign in (8.4). We continue our close look at details of Faraday's law. Equation (8.5) is just a definition of the flux of $\vec{B}$ through a surface S, and that should give us no difficulties. But in Faraday's law we can apparently use any surface S that is bounded by the loop.

We are to understand this in precisely the sane way we understand that in Ampére's law we can calculate the current through any surface bounded by the circulation path. In the case of Faraday's law, we are justified in the claim that any surface will suffice, provided it is bounded by the loop, by the fact that

$$\sum_{\text{S closed}} \vec{B} \cdot \Delta\vec{S} = 0 \tag{8.6}$$

That is, the flux of $\vec{B}$ through any *closed* surface is zero. This is nothing but the statement that there are no separable magnetic poles (see monograph Magnetostatics). Given (8.6), we can easily see why we can use any surface in (8.5), provided, of course, that surface is bounded by the loop.

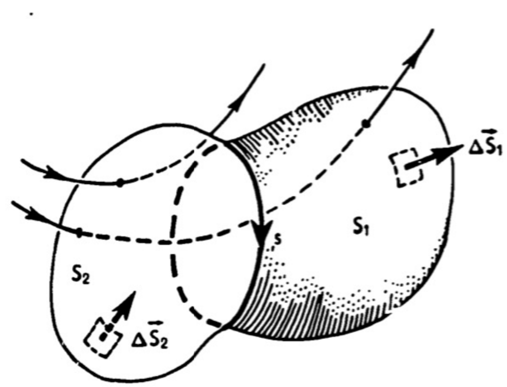


Figure 8.11: Surfaces S_1 and S_2, each bounded by the curve s, together form a closed surface. The vectors ΔS_1 and ΔS_2 are shown in the directions considered to be positive when the sense of the circulation shown about the curve s is considered positive.

That there are no separable magnetic poles means that the lines of $\vec{B}$ have no beginning and no ending. That happens sometimes because the line of t is closed, but it need not be closed. If we start to trace along a line of $\vec{B}$, we will never reach a point at which it is terminated. This means that if a line of $\vec{B}$ passes through one of the surfaces shown in Figure 8.11, it must do one of two things. It must turn around and pass back through that same surface, or it must pass through the second surface in the same direction it passed through the first. In the first case, it contributes nothing to the flux through either surface. In the second, its contribution to the flux through one surface is the same as its contribution to the flux through the other. Thus, we arc certain that the flux through

a surface bounded by the loop is the same as the flux through any other surface bounded by the same loop.

We can make the same proof in a more formal fashion. In Figure 8.11 the surfaces S_1 and S_2 form a closed surface when combined. We can use equation (8.6) on that closed surface, so that

$$\sum_{S_1 \text{ closed}} \vec{B} \cdot \Delta\vec{S}_1 - \sum_{S_2 \text{ closed}} \vec{B} \cdot \Delta\vec{S}_2 = 0$$

where we have taken due care to have the outward direction positive. From this,

$$\sum_{S_1 \text{ closed}} \vec{B} \cdot \Delta\vec{S}_1 = \sum_{S_2 \text{ closed}} \vec{B} \cdot \Delta\vec{S}_2$$

Since S_1 and S_2 here can be any two surfaces bounded by the loop, we have proved our contention.

With Faraday's law in hand we can understand the results of the experiments using the long coil shown in Figure 8.9. If the coil is tightly wound and is long enough, and if we stay near the center of the coil, then B inside the coil is very nearly the same at all points. Its magnitude is given by

$$B = \mu_0 n I$$

where n is the number of turns per unit length of the coil measured along the coil's axis, and I is the current in the coil. Further, B is directed parallel to the coil's axis. This result is derived using Ampére's law in the monograph Magnetostatics. If we are sure that the loop is oriented with its plane perpendicular to the coil's axis, then we can find the flux B threading the loop of area A. Since B is the same everywhere, then choosing S to be the plane surface of the loop, we can write

$$\begin{aligned}
\Phi_B &= \sum_S \vec{B} \cdot \Delta\vec{S} \\
&= \vec{B} \cdot \sum_S \Delta\vec{S} = \vec{B} \cdot \vec{A} \\
&= BA
\end{aligned}$$

Now putting in the magnitude of B,

$$\Phi = \mu_0 n A I$$

Then, by Faraday's law, we can say that

$$\mathcal{E} = -\mu_0 n A \frac{\Delta I}{\Delta t}$$

where $\Delta I/\Delta t$ is the time rate of change of the current in the coil. If we put this result into Ohm's law, we get

$$i = -\frac{\mu_0 n A}{R}\frac{\Delta I}{\Delta t} \tag{8.7}$$

where i is the current induced in the loop, and R is the resistance in the circuit containing the loop and the galvanometer.

All the results reported earlier for this experimental arrangement are contained in equation (8.7). If we believe equation (8.7) to be correct, then we can predict what will happen when we change the loop's area or change the resistance R. And we can predict how the result will depend upon the time rate of change of the current in the coil; in particular we would predict the behavior shown in the plots appearing in Figure 8.4 and Figure 8.5.

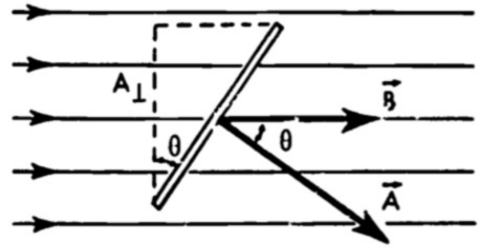


Figure 8.12: Edgewise view of a plane loop in a uniform magnetic field $\vec{B}$ when the loop is tilted. The projection of the loop's area on a plane perpendicular to $\vec{B}$ is $A_\perp$.

We can generalize equation (8.7) to take care of cases when the loop is not oriented with its plane perpendicular to $\vec{B}$. When calculating Φ_B we are really concerned with the projection of the coil's area on the plane perpendicular to $\vec{B}$ (see Figure 8.12). That projection $A_\perp$ is just equal to $A\cos\theta$; and that $\cos\theta$ factor appears when we calculate $\Phi_B = \vec{B}\cdot\vec{A}$, since the angle between $\vec{B}$ and $\vec{A}$ is also θ. Then for any orientation of the loop, (8.7) is generalized to

$$i = -\frac{\mu_0 n A\cos\theta}{R}\frac{\Delta I}{\Delta t} \tag{8.8}$$

In fact, there is nothing in our development that depends upon the *shape* of the loop; it need not be circular as we have been picturing it. Furthermore, the loop need not even lie in a plane, if we interpret to be the area enclosed by the projection of the loop onto a plane perpendicular to $\vec{B}$ (see Figure 8.13). Of course, this last statement can be true only if no line of $\vec{B}$ passes twice or more through the surface bounded by the loop.

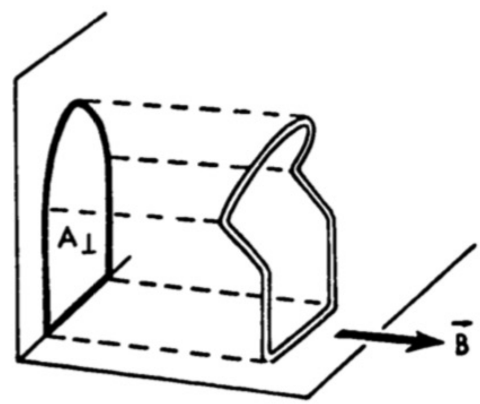


Figure 8.13: The projected area of a non planar loop. Note that this is a special case of $A_\perp$ for the loop shown.

There are cases in which the phrase "the surface bounded by the loop" might seem ambiguous. Such a case is shown in Figure 8.14. The surface we want is the one shown in Figure 8.14 (a). The surface in Figure 8.14 (b) is really not bounded at all. We can see this by noting that this surface does not have two sides; it is the famous Möbius strip. You can make a Möbius strip for yourself by

laying out a strip of paper, giving it a half twist about the long axis, and then pasting the two free ends together. Mathematics students sometimes like to plague younger brothers by asking them to color one side of this strip blue and the other red. If you have never seen this tried, you might make the effort in the secrecy of your room.

QUESTION: See Figure 8.14 (a). Assume the presence of a uniform magnetic field $\vec{B}$ directed from left to right.How would you calculate the flux Φ_B through this loop? Include a statement of any assumptions you have made to cover information not explicitly given to you.

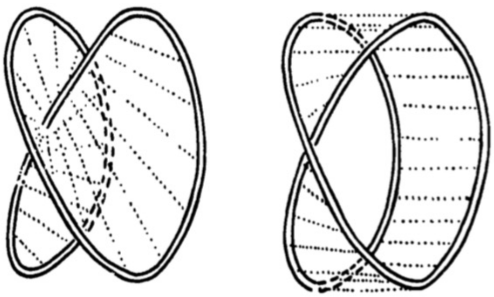

a b

Figure 8.14: Two different surfaces "bounded" by the same closed loop.

If you have thought carefully about the Question, you should be able to conclude that for the purpose of calculating Φ_B, the loop shown in Figure 8.14 (a) is identical with that in Figure 8.15 (a); and that each is equivalent to two circular loops. From this argument we would expect that had we used a loop of two turns in the experiment depicted in Figure 8.9 (a), we would have doubled the Φ_B and thus doubled the value of $\Phi_B/\Delta t$ and of the induced *emf* $\mathcal{E}$. And experimental results agree with that expectation. If we use a loop with N turns, then we would need to modify (8.8) to read

$$i = -\frac{\mu_0 N n A \cos\theta}{R}\frac{\Delta I}{\Delta t} \tag{8.9}$$

QUESTION: What is N for the loop shown in Figure 8.15 (b)? Does it matter how the ends are connected?

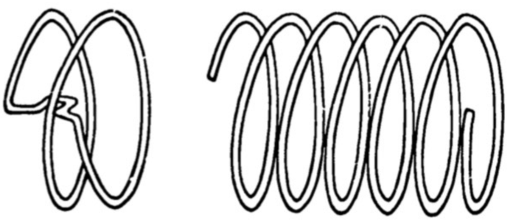

a b

Figure 8.15: (a) A closed loop of two turns. Compare with Figure 8.14 (a). (b) Now we have a coil. It is not a loop; it isn't closed.

8.3 Faraday's Law As A Circulation Law

In the case of the stationary loop an induced *emf* can appear only if there exists an electric field intensity $\vec{E}$. The *emf* in a loop is defined as being the work done on a unit charge as that charge traverses the loop once, i.e.,

$$\mathcal{E} = \sum_{s\ \text{closed}} \frac{\vec{F}}{q}\cdot\Delta s \tag{8.10}$$

from the definition of work. But what is $\vec{F}/q$ in this case? Nothing but the definition of $\vec{E}$, the electric field intensity. That means that for this case, we can write Faraday's law as closed

$$\sum_{s\ \text{closed}} \vec{E}\cdot\Delta s = -\frac{\Delta\left(\sum_S \vec{B}\cdot\Delta S\right)}{\Delta T} \tag{8.11}$$

which is a *circulation law*. Further, since the loop is stationary, all flux changes are due to changes in $\vec{B}$. In that case, we can write (8.11) as

$$\sum_{s \text{ closed}} \vec{E} \cdot \Delta s = -\sum_{S} \frac{\Delta \vec{B}}{\Delta T} \cdot \Delta S \qquad (8.12)$$

where the surface S is any surface bounded by the circulation path s – here the loop. We are forced to a startling conclusion that has enormous consequences:

> *If we have a changing magnetic field, i.e., one in which $\vec{B}$ is time dependent, then there is an associated electric field! That is, there can be an electric field even though the charge density is zero everywhere.*

Maxwell was the first to express Faraday's law of induction as a circulation law. But he generalized the law. He imagined that the law as expressed in (8.11) or (8.12) holds in empty space, so that the circulation of E need not be interpreted as the induced *emf* in a material, conducting loop. If we assume that Maxwell was correct in his assumption, and he was, then we can state Faraday's law in the following way.

> *The circulation of $\vec{E}$ about a closed path s at rest in the frame in which $\vec{B}$ is measured is equal to the negative of the time rate of change of the flux of $\vec{B}$ passing through any surface S bounded by that circulation path.*

According to this circulation law, the circulation of $\vec{E}$ is not necessarily zero in time-dependent circumstances. For time-independent cases, i.e., static cases, the circulation of E is always zero, no matter what circulation path we choose (see monograph ELECTROSTATICS). That is, static electric fields are conservative; not so for electric fields associated with changing magnetic fields. Nonetheless, the circulation law given by (8.11) is true in general, i.e., when the electric field intensity $\vec{E}$ has contributions from both charge densities (static field $\vec{E}_S$) , and changing magnetic fields (induced field $\vec{E}_i$) . If we write that $\vec{E} = \vec{E}_S + \vec{E}_i$, then the circulation of $\vec{E}$ is

$$\begin{aligned}\sum_{s \text{ closed}} \vec{E} \cdot \Delta\vec{s} &= \sum_{s \text{ closed}} (\vec{E}_S + \vec{E}_i) \cdot \Delta\vec{S} \\ &= \sum_{s \text{ closed}} \vec{E}_S \cdot \Delta\vec{s} + \sum_{s \text{ closed}} \vec{E}_i \cdot \Delta\vec{s} \\ &= \sum_{s \text{ closed}} \vec{E}_i \cdot \Delta\vec{s}\end{aligned}$$

The circulation of $\vec{E}_S$ is zero, leaving only the circulation of $\vec{E}_i$, which is equal to $\mu\Phi_B/\Delta t$ as required by the circulation law. Then we see that the circulation law is true in general, because the static contribution to the total field $\vec{E}$ does not contribute to the circulation of $\vec{E}$.

We see clearly from this argument that knowing the circulation of E$\vec{E}$does not mean that we know much about $\vec{E}$ itself. In fact, we can use this circulation law to calculate $\vec{E}$ only in special, highly symmetrical circumstances. The experimental arrangement shown in Figure 8.9 is one of these special cases. We need not consider the probe loop shown, if we accept Maxwell's generalization to empty space. Again we imagine that we are near the middle of a long, tightly wound coil, so that we may assume that $\vec{B}$ is uniform and directed parallel to the coil's axis. There are no charge densities around, so there is no static field contribution to $\vec{E}$ inside the coil.

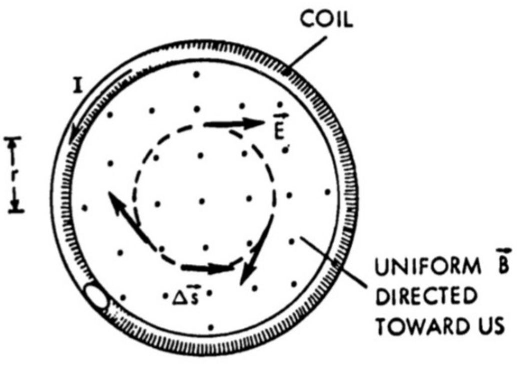


Figure 8.16: View along the axis of a long coil. With the sense of $\Delta\vec{s}$ shown, the positive flux Φ_B is directed toward us. The directions of $\vec{E}$ on the circular path are for the case of increasing Φ_B.

Figure 8.16 Is a view looking along the coil's axis. We choose as the circulation path a circle of radius r, centered on the coil's axis, and oriented so that its plane is perpendicular to the coil's axis. As shown, r is less than the inside radius of the coil. If the current in the coil is changing, then the magnitude of $\vec{B}$ is changing, and so is the flux Φ_B that passes through a surface bounded by the circulation path. Then the circulation of $\vec{E}$ around that path is not zero. Can we calculate what it is? Yes, we can use the right-hand side of the circulation law.

$$\sum_{s\text{ closed}} \vec{E}\cdot\Delta\vec{s} = -\frac{\Delta\Phi_B}{\Delta t} = -\left(\frac{\Delta B}{\Delta t}\right)\pi r^2$$

Now since we have good cylindrical symmetry, the magnitude of $\vec{E}$ at one point on the circular path must be the same as it is at every other point on the path. If there is no charge density around, then $\vec{E}$ at each point on that circle is tangent to the circle, and we can write the circulation law as

$$(E)(2\pi r) = -\pi r^2\left(\frac{\Delta B}{\Delta t}\right)$$

And now we have that

$$E = -\frac{r}{2}\left(\frac{\Delta B}{\Delta t}\right) \tag{8.13}$$

For this highly symmetric situation we have used the circulation law to get an expression for $\vec{E}$ that should be valid at all points inside

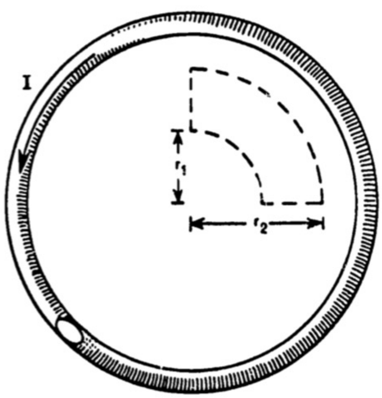
the coil and far from its ends. If, as we have assumed, $\Delta B/\Delta t$ is the same everywhere in that region, then we see that $\vec{E}$ is zero on the coil's axis and increases linearly with the distance from that axis.

Exercise

8.1 Using Gauss's law, show that there cannot be a radial component to $\vec{E}$ in the example if there are no charge densities in the vicinity.

The minus sign in (8.13) tells us the direction of $\vec{E}$. At each point on the circular path, $\vec{E}$ must have the direction just opposite to that of Δs at that point, provided that $\vec{B}$ is increasing so that $\Delta B/\Delta t$ is positive. That is the situation depicted in Figure 1.17. If the current I decreases, then $\Delta B/\Delta t$ is negative, and the direction of $\vec{E}$ is opposite to that shown in the figure.

Exercise



8.2 Using (8.13), calculate the circulation of $\vec{E}$ directly for the path shown. Arrange your result so that you can see clearly that it is equal to the negative of the time rate of change of Φ_B. Choose the case in which I is increasing.

Suppose that in the above Exercise we add a long wire along the coil's axis and let this wire carry the linear charge density λ coulomb/meter. Now calculate the circulation of $\vec{E}$ for the path shown. In what way does your result differ from that you got when no charged wire was present? Why? Does your answer to this last question depend upon the fact that the static electric field due to the line charge is radial? Explain.

QUESTION:

In either of the above two exercises, what meaning can we assign to the idea of "potential difference between two points in the field" while the current in the coil is changing?

8.4 Motional *emf*

To this point in our study of Faraday's law, we have concentrated upon *emf*'s induced in a stationary loop as a consequence of a time-varying magnetic field. In such a case, the flux Φ_B changes, because $\vec{B}$ changes with time, at least over some region of space. But there are other ways to make change for a given loop, and we are going to look at one of them now.

Consider the following simple case for which we can make some calculations. We have a piece of wire bent into a rectangle, so that it forms a closed conducting loop. We place the loop in the xy plane, as shown in Figure ?? (a). There is a magnetic field present such that $\vec{B}$ points in the direction at every point in the region of interest. Further, the magnitude of $\vec{B}$ increases with increasing y, but $\vec{B}$ at each point is constant in time.

Now we imagine that the loop is moving with the velocity v in the direction of increasing y while it remains entirely in the xy plane. The flux Φ_B threading the loop is the changing, and, if Faraday's law is correct, we should expect that an *emf* will be induced in a loop which moves in such an environment

We should realize that this is a quite different situation from that discussed in the two previous sections. This corresponds to the case shown in Figure 8.17 (b). The flux change is being created by the motion of a loop in a nonuniform, time-independent , magnetic field. We shall refer to the *emf* induced in such a fashion as a *motional emf*.

We can calculate the time rate of change of Φ_B for the simple case cited. The locations of the loop at the times t and t + At are shown in Figure 8.17 (b). If we let Φ_B be positive when it is in the $+z$ direction, we can get the corresponding $\Delta\Phi_B$ calculating the gain in Φ_B at the front edge and the loss of Φ_B at the back. That is,

$$\begin{aligned} \Delta\Phi_B &= +(B_1\, av\, \Delta t) - (B_2\, av\, \Delta t) \\ &= (B_2 - B_1)\, av\, \Delta t \end{aligned} \tag{8.14}$$

where $\vec{B}_1$ and $\vec{B}_2$ are the magnetic induction fields at the front and back edges of the loop. We have assumed that Δt is so small that $\vec{B}$ does not vary appreciably over either of the areas $(av\,\Delta t)$ at the front and back edges. Since we let $B_1 > B_2$, then $\Delta\Phi_B$ is positive; $\Delta\Phi_B$ is increasing.

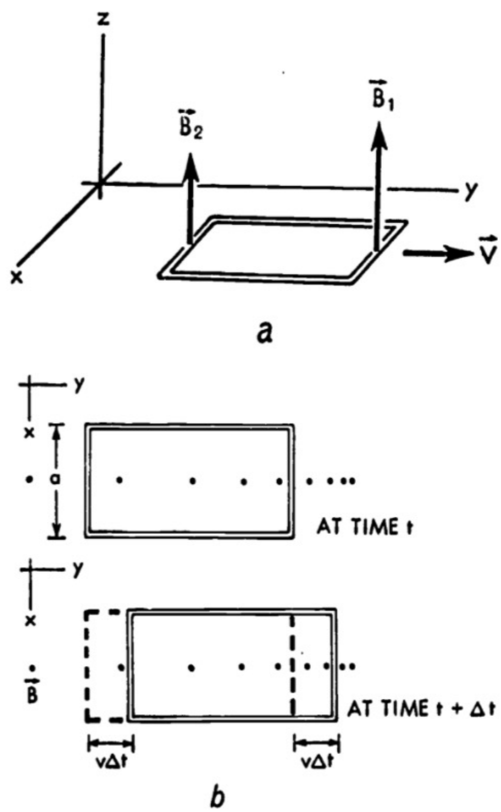


Figure 8.17: A loop moves through a nonuniform, time-independent, magnetic field. The flux Φ_B is changing.

From equation (8.14),

$$\frac{\Delta\Phi_B}{\Delta t} = (B_2 - B_1)\, av \tag{8.15}$$

so that the *emf* induced in the loop is, by Faraday's law,

$$\mathcal{E} = -(B_2 - B_1)\, av \tag{8.16}$$

and the induced current is in the clockwise sense around the loop as we view the loop in Figure 8.17 (b).

If the magnetic field is uniform, a *translation* of the loop will not produce an induced *emf*, since there will be no flux change. The quantities B_1 and B_2 in equation (??) will be equal.

But if we *rotate* the loop, even in a uniform magnetic field, then we *can* find an induced *emf* in the loop. Figure 8.18 shows a rectangular loop rotating about an axis which is perpendicular to a uniform field $\vec{B}$. In this case the flux Φ_B is changing because the projected area $A_\perp$ is changing. At the instant shown, there is an induced *emf* in the clockwise sense around the loop as viewed from the xy plane. If $\vec{B}$ is uniform, then the flux $\Phi_B$is simply $BA_\perp$ for the case shown. Then, since $\vec{B}$ is constant in time,

$$\mathcal{E} = -\frac{\Delta\Phi_B}{\Delta t} = -B\,\frac{\Delta A_\perp}{\Delta t} \tag{8.17}$$

QUESTION:

Is there an axis about which we can rotate the loop in Figure 8.18 without inducing an *emf*? If there is, does it matter what the loop's orientation is with respect to that axis?

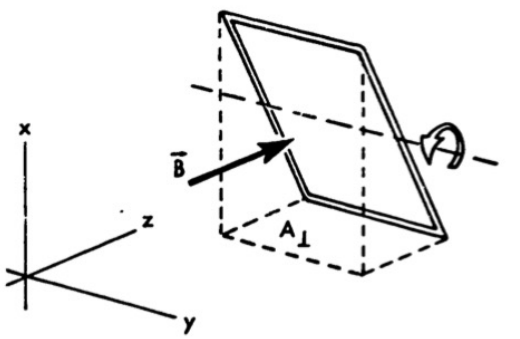


Figure 8.18: Inducing an *emf* in a loopy rotating the loop in a uniform magnetic field.

The loop in Figure 8.18 need not be a rectangle; it was drawn that way for simplicity. The loop can have any shape; it could even be a coil with many turns. In any case, we have here the fundamental concepts of the *generator*, which we shall discuss in some detail in Section 8.6.

We see now that an induced *emf* can appear if we move a loop in a time-independent magnetic field. The needed change of the flux is brought about in two ways: The loop moves through a region in which $\vec{B}$ is not uniform (Figure 8.17), and the loop's orientation relative to $\vec{B}$ changes (Figure 8.18). Of course both these could be going on at the same time; calculating $\Delta\Phi_B/\Delta t$ could become

a tricky and messy business. But it turns out that in many cases such a calculation is not difficult at all. We shall make an assertion here, and then we shall prove the assertion. There is no physics in what we shall say ; it is simply a consequence of some mathematical reasoning. And, for the moment, we shall view the assertion as an aid to calculation. The assertion: If $\vec{B}$ is independent of time, then

$$\begin{aligned} \frac{\Delta\Phi_B}{\Delta t} &\equiv \frac{\Delta}{\Delta t}\sum_S \vec{B}\cdot\Delta\vec{S} \\ &= -\sum_{s\text{ closed}} (\vec{v}\times\vec{B})\cdot\Delta\vec{s} \end{aligned} \tag{8.18}$$

where S is any surface bounded by the curve s (our loop), $\vec{v}$ is the velocity of the element $\Delta\vec{s}$, and $\vec{B}$ in the right-hand sum is the field at each $\Delta\vec{s}$. The velocity $\vec{v}$ is measured in the same frame of reference in which $\vec{B}$ is measured.

Equation (??) tells us that under certain circumstances the time rate of change of a flux can be written as a circulation. The particular circumstances for which (??) is a valid mathematical statement will appear as we develop the proof, which we now proceed to do.

Figure 8.19 shows the positions of a loop at the times t and $t+\Delta t$. At the time t, the flux is simply

$$\Phi_B(t) = \sum_S \vec{B}\cdot\Delta\vec{S}$$

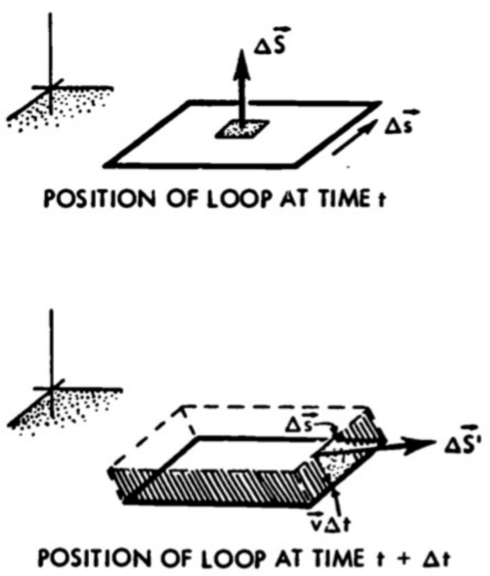


Figure 8.19: A moving loop.

To find the flux at the time $t + \Delta t$, we can use any surface bounded by the loop, provided only that the lines of $\vec{B}$ do not terminate. Or what is equivalent, provided the flux of $\vec{B}$ over any closed surface is zero. (This condition is certainly satisfied by the magnetic induction field.) Then suppose we choose as the surface at $t + \Delta t$ the original surface S plus the edge surface S' that has been generated by the moving loop. Then

$$\Phi_B(t+\Delta t) = \sum_S \vec{B}\cdot\Delta\vec{S} + \sum_{S'} \vec{B}\cdot\Delta\vec{S}'$$

If $\vec{B}$ at each point in space is the same at the time $t + \Delta t$ as it was at the time t, then the first term on the right side is just $\Phi(t)$. Then the change in the flux Φ_B that has occurred in the time Δt is

$$\Delta\Phi_B = \Phi_B(t+\Delta t) - \Phi(t) = \sum_{S'} \vec{B}\cdot\Delta\vec{S}'$$

But from Figure ??, $\Delta S' = (\vec{v}\Delta t) \times \Delta\vec{s}$ where we have been careful to keep the directions of $\Delta\vec{s}$, $\Delta\vec{S}$, and $\Delta\vec{S}'$ consistent. If we substitute $\Delta t(\vec{v} \times \Delta\vec{s})$ for $\Delta\vec{S}'$ in the sum, then we need to sum over all the $\Delta\vec{s}$ rather than over the $\Delta\vec{S}'$. Then we can write

$$\Delta\Phi_B = \Delta t \sum_{s \text{ closed}} \vec{B} \cdot (\vec{v} \times \Delta\vec{s})$$

We can see that if we are going to be able to complete our proof at all, we must be close to doing so now. Only some juggling remains. We can divide both sides by Δt to bring the left-hand side into order.

$$\frac{\Delta\Phi_B}{\Delta t} \equiv \frac{\Delta}{\Delta t} \sum_S \vec{B} \cdot \Delta\vec{S} = \sum_s (\vec{v} \times \Delta\vec{s}) \cdot \vec{B}$$

We have also changed the order of the dot product on the right-hand side, but that doesn't change anything. Comparison of what we now have with equation (8.18), what we are trying to prove, shows that we have only one thing left to do. The proof that

$$(\vec{v} \times \Delta\vec{s}) \cdot \vec{B} = -(\vec{v} \times \vec{B}) \cdot \Delta\vec{s}$$

is really rather simple, but it is a bit long. So that it will not clutter up our work here, the proof has been put at the end of this section. It is really rather nice, and you might enjoy looking at it.

With this final bit of juggling we have completed our mathematical proof. We are now prepared to state the mathematical equality given in equation (8.18). It is true for any vector field $\vec{B}$ satisfying the two conditions that we imposed in our proof. First, the flux of $\vec{B}$ over any closed surface must be zero, or, put in another way, the lines of $\vec{B}$ must have no beginning and no ending. Second, $\vec{B}$ must not change with time. It may be different at different points in space, but at any given point in space $\vec{B}$ remains constant.

Equation (8.18) contains no new physics at all. But we can apply it to the case of motional *emf*, because the two conditions on $\vec{B}$ are satisfied. The vector $\vec{B}$ always satisfies the first of these, and since we are considering loops that move in time-independent magnetic fields, $\vec{B}$ satisfies the second in this instance.

If we use equation (8.18) in Faraday's induction law, we get

$$\mathcal{E} = +\Delta t \sum_{s \text{ closed}} (\vec{v} \times \vec{B}) \cdot \Delta\vec{s} \tag{8.19}$$

where the right-hand side is just the negative of $\Delta\Phi_B/\Delta t$ for this special case.

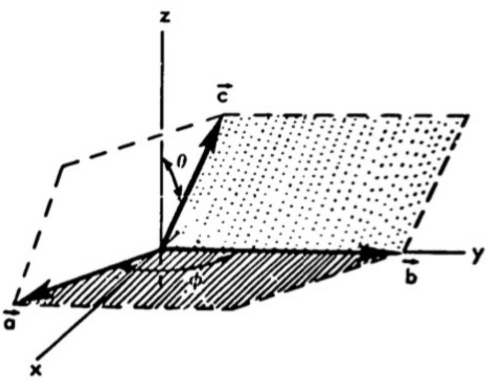
Going back over our development, we see that we did not really assume any particular shape for the loop. Equations (8.18) and (8.19) are good for any loop. Further, there is nothing in our proof of equation (8.18) requiring all parts of the loop to have the same velocity, so that equation (8.19) is valid for a loop which is rotating.

We can apply equation (8.19) to any case of motional *emf*. The right-hand side tells us to evaluate a circulation about the loop, but we know that we are really calculating the negative time rate of change of $\vec{B}$, as is required by Faraday's law.

Exercise

8.3 Use (8.19) to find the induced *emf* in the loop shown in Figure 8.17. Check your result with(8.16)

This is the proof promised earlier. Consider the three vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ shown in the diagram. For ease of reference we have put the vectors $\vec{a}$ and $\vec{b}$ in the xy plane, but nothing we do will depend upon any particular frame.



The vector product $\vec{a} \times \vec{b}$ gives a vector in the $+z$ direction (perpendicular to the plane containing $\vec{a}$ and$\vec{\ }$. The magnitude of this vector, $ab \sin \Phi$ is just the area of the parallelogram with sides $\vec{a}$ and $\vec{b}$. The scalar product of $\vec{a} \times \vec{b}$ and $\vec{c}$ gives the volume of the parallelepiped with edges $\vec{a}$, $\vec{b}$, and $\vec{c}$. That is, it gives the area of the base times the vertical height $c \cos \theta$.

$$\text{Volume} = (\vec{a} \times \vec{b}) \cdot \vec{c}$$

If we take the vector product the other way around, i.e., as $\vec{b} \times \vec{a}$, then we get a vector in the $-z$ direction. The scalar product of that vector with $\vec{c}$ gives a negative number, just the negative of the volume. So if we want the volume to come out as a positive quantity, we need the cross product in the order $\vec{a} \times \vec{b}$. Then the angle between the vectors $\vec{a} \times \vec{b}$ and $\vec{c}$ is less than $\pi/2$.

Of course it cannot matter which face of the parallelepiped we choose to be the base. We have used the one defined by $\vec{a}$ and $\vec{b}$, but we could as well start with the one defined by $\vec{b}$ and $\vec{c}$ or the one defined by $\vec{c}$ and $\vec{a}$.

We just have to keep track of the order in which we take each vector product, so that we are sure we get a positive result for the volume. Then it must be that

$$(\vec{a} \times \vec{b}) \cdot \vec{c} = (\vec{b} \times \vec{c}) \cdot \vec{a} = (\vec{c} \times \vec{a}) \cdot \vec{b}$$

We have shown this result for a particular set of vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$, but a little thought should convince you that the conclusion is valid for any set. If the vectors we choose are such that $(\vec{a} \times \vec{b}) \cdot \vec{c}$ is positive, then the other two arrangements will be positive too. If $(\vec{a} \times \vec{b}) \cdot \vec{c}$ is negative, then the other two arrangements will also be negative. The result we have here is independent of the labeling and relative orientation of the three vectors.

Now we use the first and third of these, reversing the order of the vector product in the third and putting in a minus sign to take care of that reversal

$$(\vec{a} \times \vec{b}) \cdot \vec{c} = -(\vec{a} \times \vec{c}) \cdot \vec{b}$$

Since $\vec{a}$, $\vec{b}$, and $\vec{c}$ are any vectors, this completes the proof that

$$(\vec{v} \times \Delta\vec{s}) \cdot \vec{B} = -(\vec{v} \times \vec{B}) \cdot \Delta\vec{s}$$

which is what we asserted earlier.

8.5 Magnetic Force On A Moving Charge

Something rather interesting happened in the last section, but we did not pay any attention to it at the time. Now we want to take a closer look at equation (8.19).

$$\mathcal{E} = \sum_{s \text{ closed}} (\vec{v} \times \vec{B}) \cdot \Delta\vec{s} \qquad (8.19)$$

We got the right-hand side of this by seeing what $-\Delta\Phi_B/\Delta t$ was for a special case: a loop moving in a time-independent magnetic field. That is, equation (8.19) is just Faraday's law for the special case where $\mathcal{E}$ then is the resulting motional *emf*.

Now the *emf* is the work done on a unit charge as that charge traverses the circuit one time. But what is the force doing that work in our special case? There are no charge densities around to give a static electric field. And in the frame in which $\vec{v}$ is measured the magnetic induction field $\vec{B}$ is not changing with time; i.e., $\Delta\vec{B}/\Delta t =$

0, so that in that frame there is no induced electric field either. Then how shall we account for the resulting *emf*?

Using the definition of $\mathcal{E}$ in mathematical terms, i.e.

$$\mathcal{E} = \sum_{s\text{ closed}} \left(\frac{\vec{F}}{q}\right) \cdot \Delta\vec{s} \tag{8.20}$$

We can write equation (8.19) as

$$\sum_{s\text{ closed}} \left(\frac{\vec{F}}{q}\right) \cdot \Delta\vec{s} = \sum_{s\text{ closed}} (\vec{v} \times \vec{B}) \cdot \Delta\vec{s} \tag{8.21}$$

where both sums are over the same closed loop.

There is no logical or mathematical basis for equating the bracketed terms on the two sides of equation (8.19). That is because it might be that

$$\frac{\vec{F}}{q} = \vec{v} \times \vec{B} + \vec{C}$$

where $\vec{C}$ is a vector field whose circulation is always zero, as, for example, in the static electric field. But if for our special case we do set the bracketed terms equal to each other, i.e., if we just set $\vec{F}/q$ equal to $\vec{v} \times \vec{B}$ at every point along the loop, then we get that

$$\vec{F} = q\,(\vec{v} \times \vec{B}) \tag{8.22}$$

at every point on the loop. But this is just the equation for the magnetic force on a moving charged particle (see monograph Magnetostatics).

Suppose we move a conductor in a magnetic field. There are charges that are free to move inside the conductor, and they do so move when they experience the $q\,(\vec{v}\times\vec{B})$ force. In a conductor such as copper, electrons are the particles that can move around. When a copper wire is given a velocity $\vec{v}$, the electrons have that velocity too. Then the electrons experience the $q\,(\vec{v} \times \vec{B})$ force that may result in an *emf* in the loop. Figure 8.20 is meant to illustrate what happens. In the case shown, the $\vec{v} \times \vec{B}$ force is along the wire, and it can thus contribute to an *emf*. Since q is negative for the electron, the force $\vec{F}$ is directed opposite to $\vec{v} \times \vec{B}$. The conventional current I, however, is still in the direction of $\vec{v} \times \vec{B}$.

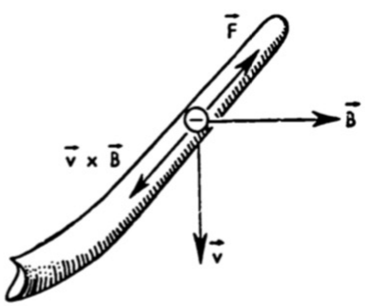


Figure 8.20: A segment of a conducting wire which has the velocity $\vec{v}$ while moving in the field $\vec{B}$. $\vec{F}$ is the force on an electron in the wire.

If we try to trace the motion of one of these electrons, if we try to keep track of the forces on it as it moves inside the conductor, we

can get ourselves into a great tangle. As soon as the electron gains a component of velocity along the wire, then the force it feels is no longer that shown in Figure 8.20. That is because its velocity is no longer $\vec{v}$. The velocity $\vec{v}$ appearing in equation (8.19), and thus in equation (8.22), is the *velocity of the wire*, not the velocity of the electron when the electron travels inside that wire. Further, we certainly know that the electron experiences an enormously complex force field as it travels through the conductor. Nevertheless, if we assume the simple force law given by (8.22), we come up with the right result for the *emf*. That certainly seems strange, but it is true.

With the interpretation of $\vec{v} \times \vec{B}$ as a force per unit charge, the right side of (8.19) takes on a new physical meaning. We need not think of it as being the negative of the time rate of change of which it certainly is. We can think of it as being the direct calculation of the *emf* using the defining (8.20) with $\vec{v} \times \vec{B}$ being the force per unit charge.

All of this leads us to the following point of view: A charged particle can experience a force that is *velocity independent*, and we write this force as $q\vec{E}$ The electric field intensity $\vec{E}$ can be the consequence of a distribution of charge density or of a time—varying magnetic field. A charged particle can also experience a *velocity-dependent* force $q\,(\vec{v} \times \vec{B})$, if there is a magnetic field present. If we put these together, we can say in general that the electromagnetic force on a particle with charge q is

$$\vec{F} = q\,(\vec{E} + \vec{v} \times \vec{B}) \tag{8.23}$$

This has come to be called the *Lorentz force*. Its application to the motions of free particles in empty space is often more straightforward than it is when applied to particles which themselves move around in moving materials. Since $\vec{v}$ is the velocity of the material, not that of the particle inside the material, it seems as if the particle is in a *field* $\vec{v} \times \vec{B}$ that exists inside materials whenever they move in magnetic fields.

The strange nature of a velocity-dependent force is discussed in a later section. Sooner or later, we must look into the difficulties encountered when we try to reconcile the concept of a velocity dependent force field with the principle that all inertial frames of reference are equivalent.

8.6 Generators And Motors

Modern electrical technology began with Faraday's discovery, for then engineers had the ideas they needed to permit the design of machines to create, deliver, and use electric energy. When they were restricted to batteries, electric energy was so limited that engineers were unable to make much use of it. Anyway, batteries run down. Now electricity can light cities, turn millions of wheels, heat homes, and carve mountains – all because Michael Faraday wondered about the ways that magnets and currents were related.

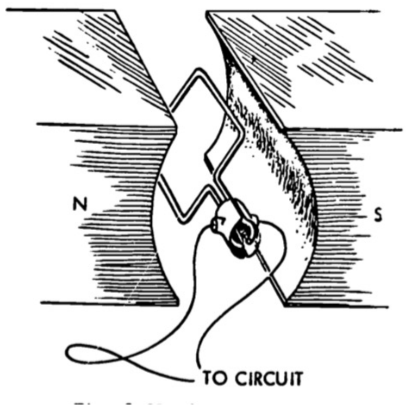


Figure 8.21: A generator.

Faraday had to suffer officials who visited his laboratory while he was working, and he was often called upon to give public lectures on his work. There is no way to know how many times he was asked, "What's the use of all this?" But he was ready with a reply when Gladstone, then Chancellor of the Exchequer, interrupted him impatiently, "But, after all, what use is it?" Faraday fired back, "Sir, you may one day be able to tax it." Such opportunities come rarely, even to men of Faraday's stature.

Benjamin Franklin responded to questions like that with a question of his own. "What is the use of a baby?" What exactly is this baby that grew up to be taxed? Its essence is in Figure 8.18 which is meant to show a loop rotating in a magnetic field

Suppose we arrange, by some means or other, to keep a loop rotating in a magnetic field. And suppose, too, that we arrange to make this loop part of a larger electric circuit. Figure 8.21 shows what we have in mind, although it doesn't show how we intend to sup- port the loop or to keep it rotating. But while that loop is rotating, it is a source of *emf* for the circuit, and, using equation (8.19), we should be able to calculate that *emf*.

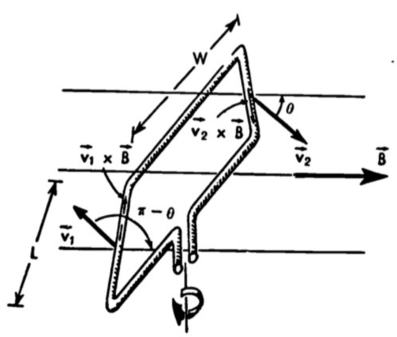


Figure 8.22: Finding $\sum_{s\ \text{closed}} (\vec{v} \times \vec{B}) \cdot \Delta\vec{s}$ for a rotating loop.

A look at Figure 8.22 should help us to calculate the circulation of $\vec{v} \times \vec{B}$ around the rotating coil. If the axis of rotation of the loop passes through the loop's center and is parallel to two of the edges, then each of these edges has the same speed v. In terms of the notation on the figure, $v_1 = v_2$ further, if we have a uniform magnetic field, then the magnitude of $\vec{v}_1 \times \vec{B}$ is equal to the magnitude of $\vec{v}_2 \times \vec{B}$ since $\sin(\pi - \theta) = \sin\theta$. If we take the sense of the circulation about the loop to be along the directions of both $\vec{v}_1 \times \vec{B}$ and $\vec{v}_2 \times \vec{B}$, then these two sides of the loop contribute

$$2vBL \sin\theta$$

to the circulation.

The two other sides contribute nothing at all to the circulation, because at every point on them $\vec{v} \times \vec{B}$ is perpendicular to $\Delta\vec{s}$. Then we already have the circulation of $\vec{v}_1 \times \vec{B}$ around the loop, so we can say that the *emf* induced in the loop is

$$\mathcal{E} = 2vBL \sin\theta \tag{8.24}$$

where we remember that v is the speed of those sides which are always perpendicular to $\vec{B}$. If we introduce the other dimension of the loop; i.e., w, we see that each point on these two sides travels in a circle of radius $w/2$. Then we can write that $v = (w/2)\omega$, where ω is the angular velocity of the loop; i.e., ω is just $\Delta\vec{B}/\Delta t$ measured in radians/ second . With this substitution, equation (8.24) becomes

$$\mathcal{E} = BA\omega \sin\theta \tag{8.25}$$

where we have substituted A, the loop's area, for the product wL.

We want another calculation on hand to help in our discussion of the physical content of equation (8.25). When the loop is in the position shown in Figure 8.22, then the flux Φ_B that the loop intercepts is

$$\Phi_B = BA \cos\theta \tag{8.26}$$

At the instant shown in Figure 8.22, the flux Φ_B is positive but it is decreasing. When Φ_B is decreasing, then $\Delta\Phi_B/\Delta t$ is negative. Since a minus sign appears in Faraday's law itself, then the *emf* should be positive at that instant.

Figure 8.23 shows all the pertinent relationships for the rotating loop over one complete rotation. Figure 8.23 (a) displays the physical position of the loop, the second position corresponding roughly to that shown in Figure 8.22. For the plots in Figure 8.23, we have let θ be zero at $t = 0$, so that for constant angular velocity , ω we have that $\theta = \omega t$.

Comparing the plots in (b) and (c) of Figure 8.23, we see that $\mathcal{E}$ has a maximum magnitude when Φ_B is zero, and $\mathcal{E}$ is zero when Φ_B has its maximum magnitude. Well, that is really what we expect, since Faraday's law says that the induced *emf* is proportional to the time rate of change of Φ_B. And that time rate of change $\Delta\Phi_B/\Delta t$ is just proportional to the slope of the Φ_B vs. θ curve in Figure 8.23 (b). And certainly that slope has its maximum magnitude at $\theta = \pi/2$ and $\theta = 3\pi/2$, the same place that $\mathcal{E}$ has *its* maximum magnitude. Since $\Delta\Phi_B/\Delta t$ is negative at $\theta = \pi/2$, then $\mathcal{E}$ is positive there. Since $\Delta\Phi_B/\Delta t$ is positive at $\theta = 3\pi/2$, then $\mathcal{E}$ is negative there. And since $\Delta\Phi_B/\Delta t$ is zero at $\mathcal{E} = 0$ and $\theta = \pi$ then $\mathcal{E}$ is zero at those value of θ.

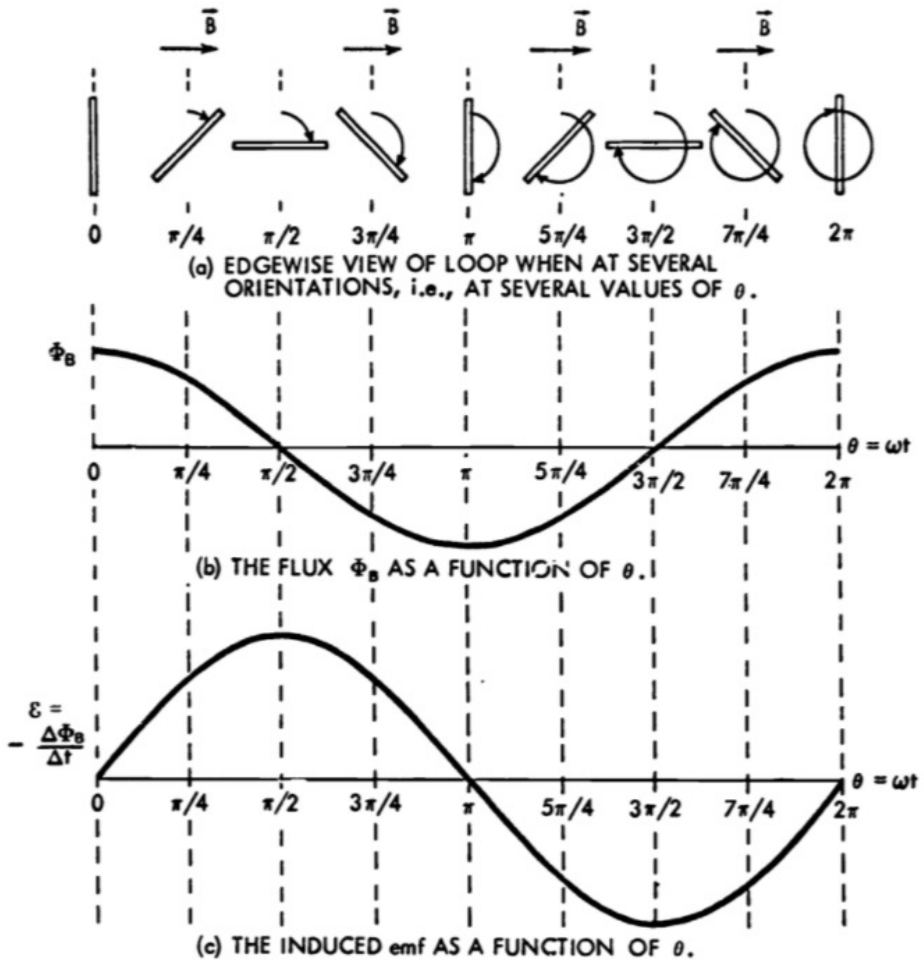


Figure 8.23: A plane loop rotates with constant angular velocity in an uniform magnetic field.

Exercise

8.4 Using equations (8.25) and (8.26) along with Faraday's law, convince yourself that

$$\frac{\Delta \cos(\omega\, t)}{\Delta t} = -\,\omega \sin(\omega\, t)$$

where ω is a constant.

Figure 8.23 tells us that the *emf* in the loop is in one sense when $0 < \theta < \pi$ and in the opposite sense when $\pi < \theta < 2\pi$. But the contact that rotating loop makes with the circuit can be arranged so that the *emf* in the circuit is always in the same direction. That sort of contact is shown in Figure 8.21; it is called a split-ring *commutator*. If the commutator is arranged as showing in Figure 8.24, then the

emf in the circuit is in one sense for half a cycle of the loop, and in the opposite sense for the other half cycle. In that case, we say that we have an a-c (alternating current) generator.

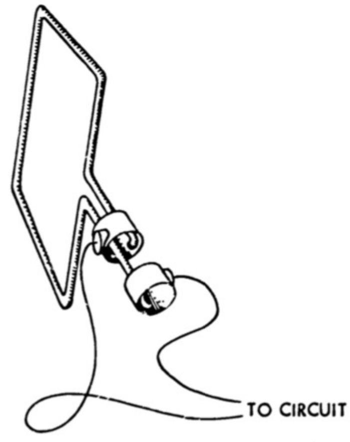


Figure 8.24: Electrical connection for an a-c generator.

Exercise

8.5 Plot the *emf* in the complete circuit when a split-ring commutator is used such as is illustrated in Figure 8.21. At what values of θ do the contacts pass through the "splits"? Why?

Now we have the fundamental principle of a generator. There are certainly refinements that we could make. For instance, we could use a loop of many turns, so that the *emf* would be increased provided we can maintain a sufficiently high angular velocity. But we still have an important question to answer. How do we keep the loop rotating? We need an answer, because we get an *emf* only when the loop rotates.

One way to crank this system is to attach a paddle wheel to the loop and then put it at the bottom of a waterfall . The water is going to lose all that potential energy anyway, so it may as well fall on the paddle wheel and, thus, spin the loop. The large, specially designed, paddle wheels located at places such as Niagara Falls are called turbines. If there is no nearby waterfall, someone might build a high dam on a river, so that giant pipes running down the inside of the dam provide an artificial waterfall. Lacking a dam, the turbine can be driven by steam at high pressure. In that case, we need to boil water, which requires large quantities of coal or else a nuclear power reactor. In any case, the goal is to rotate a turbine.

Engineers have spent a lot of time designing electric power plants. And they have solved an enormous number of complex technical problems so that they can operate these plants at the highest efficiency. Our short description does not do justice to what they have accomplished. We have just looked at the basic scientific law that is the heart of the matter.

If the electrical energy developed at a plant is to be delivered to some place that is far away, there are lots of other interesting problems to solve. But we cannot go into the transmission problems here. We are going to see what we can do with this electric energy once it has been delivered to us. We are normally provided with two electrical contacts; in houses the usual potential difference between one of these contacts and the other alternates, taking on all values

between about +155 volts and −155 volts, the root mean square (rms) value being about 110 volts. If you want to run washing machine and drier you will likely need a pair at 220 volts rms potential difference. A much higher potential difference is maintained in transmission lines, because energy losses turn out to be less along the way when the voltage is high. But the voltage is reduced by a series of transformers that are located between the transmission lines and our houses.

What can we do with this potential difference? Well, we can do some obvious things. We can put that potential difference across some resistance so that we get heat. The resistance can be in thin wire embedded in ceiling plaster; that will heat a room. Or the resistance can be in a coil on the top of an electric range; that will boil water and cook food. If the resistance is in the filament of a light bulb, then we can heat that filament; that will light our rooms.

Of course, we can use that potential difference to run a radio or television transmitter and to activate radio and television receivers. We shall not go into the modern technology of electronics that has become so important. We are going to leave out all those marvelous gadgets used for communications, for calculations that go on inside a computer, and for detecting the presence of subatomic particles. We are going to study something that seems much more prosaic: the electric motor.

Without the electric motor, the industrial revolution would surely have fizzled. The electric motor does a very large fraction of the work that needs doing in an industrialized, technological society. And anyway, the electric motor is easy to understand, once we understand the electric generator. The electric motor is just the electric generator driven backward. Even though true, that doesn't tell us much, so we want to look a little closer at the way such a motor works.

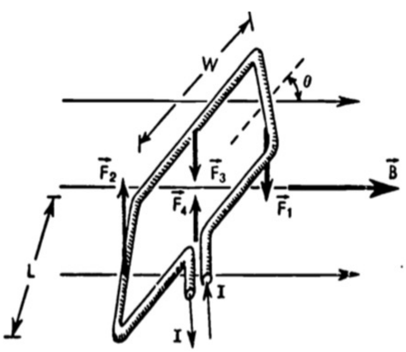


Figure 8.25: Forces on a current-carying loop in a uniform magnetic field.

Suppose we have a loop sitting still in a magnetic field. Figure 8.21, used in the discussion of a generator, pictures the situation. But now suppose that the loop is part of an electric circuit that has a source of *emf* in it, a battery or a generator. As soon as we complete that circuit by closing a switch, current appears in the loop. According to Ampere's law, a current- carrying wire experiences a force in the presence of a magnetic field. Let's look at the forces on that loop.

The force on a current element $I\Delta\vec{s}$ is just $I\Delta\vec{s} \times \vec{B}$, so we need to sum the forces on all the elements in the loop. In Figure 8.25, $\vec{F}_1$ a is the force on one side of the loop, and its magnitude is just IBL.

The force $\vec{F}_1$ on the opposite side has the same magnitude. Since these two forces are in opposite directions, their sum is zero.

The force $\vec{F}_3$ has the magnitude $IBw \sin\theta$, and so does the force $\vec{F}_4$. Therefore $\vec{F}_3 + \vec{F}_4 = 0$, since they, too, are oppositely directed. We have ignored the little gap in side 4, but we can certainly make that gap as small as we want.

Then the net force on the loop is zero. But it certainly won't stay at rest. There is a net torque on the loop created by the forces $\vec{F}_1$ and $\vec{F}_2$. The magnitude of that torque is

$$\tau = IBLw \cos\theta$$

We can write this in vector form

$$\tau = I\vec{A} \times \vec{B} \tag{8.27}$$

where $\vec{A}$ is perpendicular to the plane of the loop and has the magnitude Lw, the loop's area. We have also made use of the fact that the angle between $\vec{A}$ and $\vec{B}$ is $(\pi/2 + \theta)$ and that $\cos\theta = \sin(\pi/2 + \theta)$.

Now we see that there is a torque on the loop, so that the loop will have an angular acceleration. If we connected a pulley to the loop, that torque would turn the pulley and lift a weight hanging from it. That is exactly what an electric motor does: It turns a shaft to which pulleys or other devices can be attached. So it is true: An electric motor is just a generator operated backward. If we run a current through a loop that sits in a magnetic field, we get a torque on the loop. And we can use that torque to do work.

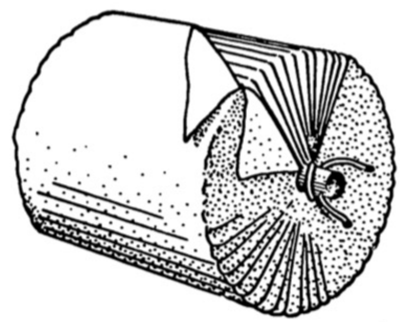


Figure 8.26: A motor loop wound to give a nearly constant torque.

We have described a very rudimentary motor. It will operate better if we wind the loop as shown in Figure 8.26. Then the torque is very nearly constant and it is always greater than the torque on a loop of a single turn. Most motors have more refinements, but they all operate on the principle that tells us that a magnetic field produces a force on a current-carrying wire.

Faraday's law determines the way a generator works and Ampére's law determines the way a motor works. It is all wrapped up in Figure 8.27. Faraday's baby, now full grown, has changed the face of the planet.

8.7 The Betatron

Most of the practical applications of induced *emf* are instances of motional *emf* rather than examples of induced *emf* as the result of a

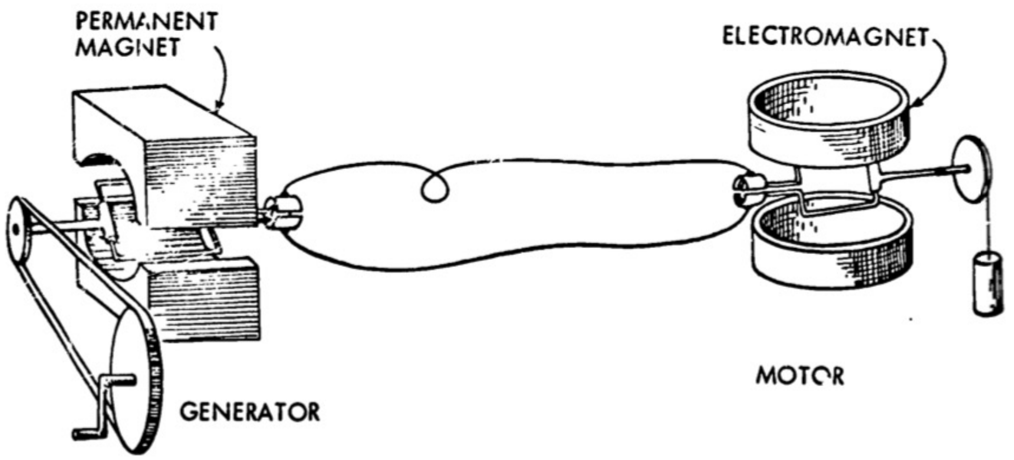


Figure 8.27: Crank the handle to spin the generator loop to create an *emf* to send a current through the motor loop to establish a torque to lift the weight.

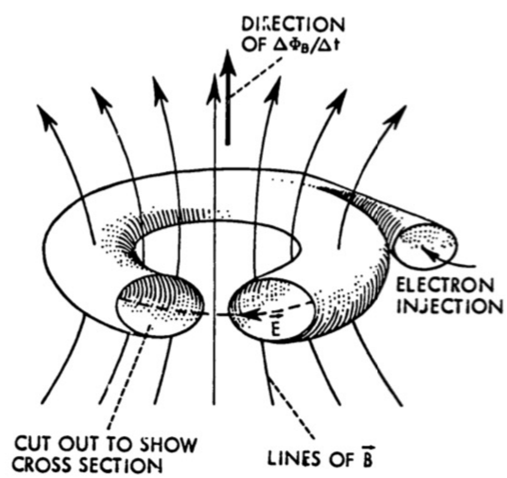


Figure 8.28: The doughnut of a betatron.

time-varying magnetic field. That is because it is easier to move a loop in a controlled way than it is to change continuously a magnetic field. But there are uses for the time-varying field approach, and one of the more striking is in the operation of a betatron.

The betatron is a machine designed to accelerate electrons to quite high velocities. Electrons are often called beta particles when they are not bound to atoms, thus the name betatron. All accelerators, including betatrons, have lots of electronic gear attached to them. There are many technical problems that must be solved before a betatron will perform satisfactorily, but we are going to ignore most of the problems and concentrate on the way a changing magnetic field is used to induce an electric field which in turn accelerates the electrons.

The heart of a betatron is a hollow toroidal affair usually made of ceramic. Figure 8.28 is a schematic picture of one of these. You can see why those in the business call this the doughnut. We pass over a lot of hard work and assume that we can evacuate the doughnut and that we have arranged to feed free electrons into it. Also, we somehow supply a magnetic field which is directed perpendicular to the doughnut's horizontal median plane at all points on that plane and which s cylindrically symmetric about the axis of the doughnut. Further, we arrange to have that field be variable in time.

Suppose we have an electron moving inside the doughnut in the median plane and in a circular orbit of radius r concentric with the axis of the doughnut, that radius being determined by the electron's velocity $\vec{v}$ and the magnetic induction field $\vec{B}_0$ at the

electron's position. Now we suddenly increase the magnetic field, so that there is a $\Delta\Phi_B/\Delta t$ through the electron's orbit. We can use Faraday's law in empty space, as Maxwell suggested, so that if we take the circulation path to be the electron's orbit,

$$2\pi r E = -\frac{\Delta\Phi_B}{\Delta t} \tag{8.28}$$

since the magnetic induction field is cylindrically symmetric about the doughnut's axis. The induced electric field $\vec{E}$ at the orbit is tangent to the orbit, so that if we have the electron moving in the proper direction to begin with, the force $q\vec{E}$ will increase its speed. We can use equation (8.28) along with Newton's second law of motion to give so that if we take the circulation path to be the electron's orbit,

$$-eE = +\frac{e}{2\pi r}\frac{\Delta\Phi_B}{\Delta t} = \frac{\Delta(mv)}{\Delta t}$$

where $m\vec{v}$ is the electron's linear momentum, and $q = -e$ is the electron's charge. From the last two terms, we have

$$\Delta(mv) = \frac{e}{2\pi r}\,\Delta\Phi_B \tag{8.29}$$

where $\Delta(mv)$ is the change in the electron's momentum that occurs over the time it takes to change the flux by $\Delta\Phi_B$.

In getting to equation (8.29), we have assumed that the electron continues to move in the same circle of radius r as the flux increases. How can we arrange to accomplish that? Using Newton's law again and the fact that the electron experiences the centripetal force $-e(\vec{v} \times \vec{B})$, we get

$$evB_0 eE = \frac{mv^2}{r}$$

where v^2/r is the electron's centripetal acceleration, and B_0 is the magnetic induction field at the electron's position; i.e., at the circular orbit. From this, we get $mv = erB_0$ and thus

$$\Delta(mv) = erB_0 \tag{8.30}$$

where ΔB_0 is the change in the magnetic induction field at r that accompanies the change in the electron's momentum $\Delta(mv)$.

The momentum change in equation (8.29) is the same as that in equation (8.30), so it must be that

$$\Delta B_0 = \frac{1}{2}\frac{\Delta\Phi_B}{\pi r^2} \tag{8.31}$$

We arranged the terms in this way because we can say that

$$\Delta\Phi_B = \Delta(\pi r^2 B_{av}) = \pi r^2 \Delta B_{av}$$

where B_{av} is the average magnetic induction field that exists over the area of the circle. If we put this into equation (8.31), the πr^2 terms cancel and we are left with

$$\Delta B_0 = \frac{1}{2}\Delta\Phi_B\, B_{av} \tag{8.32}$$

Then the condition for keeping the electron in a circular orbit is simply that the change in $\vec{B}$ at the orbit must equal just one-half the change in the average $\vec{B}$ over the area ringed by that orbit. One way to accomplish that is to have B_0 itself equal to half of B_{av} at all times, but there are certainly other ways.

We have not described how we go about assuring that the electrons do not wander away from the median plane, or how to synchronize the injection of electrons into orbit with the changing magnetic fields, nor do we intend to. The arts of designing, building, and operating particle accelerators are complex and mysterious for the uninitiated.

But we have seen how two very important ideas are used in the betatron. When there is a time-dependent magnetic field, then the circulation of $\vec{E}$ can be different from zero, even in empty space. A time-varying magnetic field induces an electric field, and that electric field is just what accelerates the electron in its orbit. Also, when a charged particle has a velocity in a magnetic field, then it experiences the force $q\,(\vec{v} \times \vec{B})$. And that force is just what keeps the electron in the circular orbit. In short, we need to use the full Lorentz force $q\,(\vec{E} + \vec{v} \times \vec{B})$ to describe the appropriate behavior of electrons in a betatron.

Chapter 9

Modification Of Ampére's Law

9.1 Ampére's Circulation Law For Steady Currents

Ampére's circulation law for steady currents is developed and explained in the monograph Magnetostatics. The law was applied only to cases in which the circulation path is entirely in empty space, and we shall continue that restriction here.

The law says that the circulation of the magnetic induction field $\vec{B}$ about any closed path is proportional to the current I passing through any surface bounded by that path. We can express this statement in mathematical terms:

$$\sum_{s \text{ closed}} \vec{B} \cdot \Delta\vec{s} = \mu_0 I \tag{9.1}$$

The current I passing through a surface is just the flux of the current density $\vec{j}$ through that surface, i.e.,

$$I = \sum_{S} \vec{j} \cdot \Delta\vec{S} \tag{9.2}$$

If the circulation path over which we sum $\vec{B} \cdot \Delta\vec{s}$ does not encircle a current, then that circulation is zero. This does not imply that $\vec{B}$ itself is zero at all the points on such a path; it implies only that the positive and negative contributions to the circulation cancel exactly.

Figure 9.1 shows a few kinds of circulation paths and two sorts of surfaces bounded by circulation paths. These serve as reminders

of the way Ampére's circulation law works. You should pay particular attention to the relationship between the sense in which the circulation path is traversed and the direction in which the current is considered positive. This convention is the same as the one we used in Faraday's law when connecting the sense of the circulation of $\vec{E}$ with the positive direction of $\Delta\Phi_B/\Delta t$. So far as equations (9.1) and (9.2) are concerned, this is the convention relating the sense of $\Delta\vec{s}$ with positive direction of $\Delta\vec{S}$; the same one we used throughout Chapter 8 and which was illustrated in Figure 8.4.

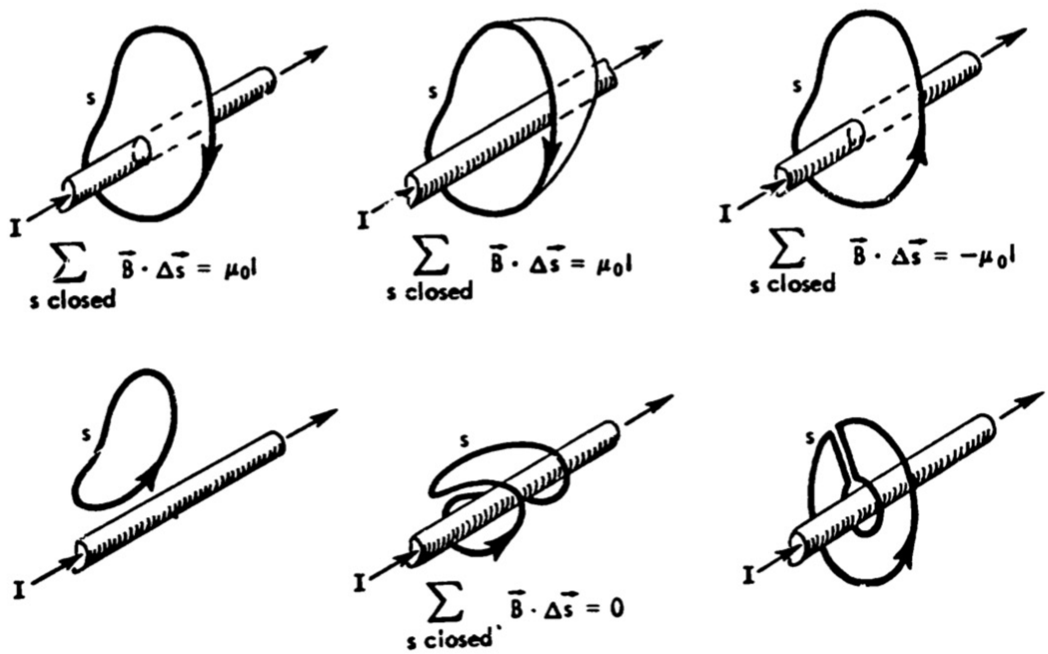


Figure 9.1: Reminders of the way Ampere's circulation law works.

There is a simple way to decide when a closed path is such that the circulation of $\vec{B}$ around it will be zero. While we understand that currents are not always in wires, assume for the moment that the current is in a wire as it is shown in Figure 9.1. Imagine now that you have taken a length of string, laid it along the circulation path, and knotted the loose ends. Now, in your imagination, of course, if you could pull the string away from the wire, then the circulation of $\vec{B}$ around that path will be zero. In that case, we would say that the path, or the siring, did not encircle the current. You might check this with the cases shown in Figure 9.1.

From the discussions in the monograph Magnetostatics, we are

convinced that Ampére's circulation law is valid for time-independent situations, i.e., when the currents are steady and the charge densities are constant. What about time-dependent situations? Will the law be valid when current or charge densities are changing? It turns out that the law is not applicable in such cases, at least not as it stands in equation (9.1). In the next section we are going to look at a rather common situation for which Ampére's circulation law certainly does not work.

9.2 A Time-Dependent Situation

As an example of a time-dependent situation, we are going to investigate the state of affairs depicted in Figure 9.2 (a). We suppose that we are charging a capacitor consisting of two conducting plates that are circular and parallel. The figure shows the long straight wires leading to these plates, and we assume that the rest of the circuit is so far away that it does not affect what goes on in the region we are investigating.

This is a time-dependent situation. We know that at least one thing is changing – the charge on the plates. We shall find a fundamental contradiction if we apply Ampére's law here. That contradiction will convince us that Ampére's circulation law is *not* generally applicable to time-dependent situations.

Let's see how the law behaves. Suppose we pick the circulation path s which is a circle of radius r and which is concentric with the current-carrying wire. That path is shown in Figure 9.2 (b). We can choose the plane surface bounded by s to be our surface S, and that is also shown in Figure 9.2 (b).

Since we have a nice symmetrical arrangement, we are quite sure that at any instant the magnitude of $\vec{B}$ is the same at every point on the circular path s. If we assume further that the angle between $\vec{B}$ and the circle's tangent is the same at every point, then we can calculate the circulation that appears on the left-hand side of Ampére's law. We get

$$\sum_{s \text{ closed}} \vec{B} \cdot \Delta\vec{s} = 2\pi r B_t \tag{9.3}$$

where B_t is the component of $\vec{B}$ tangent to the circle, assumed to be the same everywhere on the circular path. We have some confidence in this result, although we should know what the currents are like *in the capacitor plates* before we give it too much credence. For

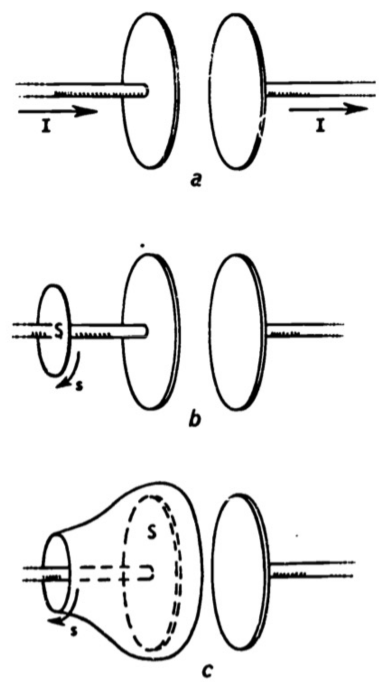


Figure 9.2: A charging capacitor, a time-dependent situation.

our purposes here, the detailed correctness of equation (9.3) is not important. Anyone too disturbed by the many assumptions made in reaching this result can just substitute the average value $\bar{B}_t$ for B_t, and then the result will be true by definition.

We can calculate the right-hand side of Ampére's law too. Using the plane surface shown in Figure 9.2 (b), we get

$$\mu_0 \sum_S \vec{j} \cdot \Delta\vec{S} = \mu_0 I \tag{9.4}$$

where I is the current in the wire at the same instant at which we calculated the circulation to get equation (9.3).

If we were to equate the results in equations (9.3) and (9.4), as we can when Ampére's law is correct, then we would get a result for B_t that looks like what we got for an unbroken, infinitely long, straight current (see Magnetostatics). And we would expect that result to be reasonably good, particularly if the circulation path is far away from the capacitor or if the capacitor plates are small and close together.

Of course, Ampére's law tells us that we may choose any surface bounded by s, so we could have chosen the vase-like one shown in Figure 9.2 (c). That one is also bounded by s, but it passes between the capacitor plates. Using that S to calculate the right-hand side of Ampére's law, we get

$$\mu_0 \sum_S \vec{j} \cdot \Delta\vec{S} = 0 \tag{9.5}$$

since no charge at all passes through that surface; i.e., the current density $\vec{j}$ is zero everywhere on that surface.

Now we can see the contradiction. We certainly cannot equate equation (9.3) to equation (9.4) and then turn around and equate equation (9.3) to equation (9.5) too. Since $\vec{B}$ cannot be both not zero and zero at the same time, we can expect trouble when we try to use Ampére's law in time-dependent circumstances. Of course, no one ever told us that we could use the law when there are time variations in currents or charge densities. We just wanted to try extending the range of application of the law, and we were not successful.

If we persist in trying to extend the applicability of Ampére's law, we shall need to be more careful – and thoughtful. Maybe we should first understand why the law does not work in time-dependent situations. That is what we look into in the following section.

9.3 Charge Conservation And Ampére's Law

We are going to see why Ampére's circulation law fails in time-dependent situations. Once we understand that, we can investigate the possibilities of generalizing the law so that it will work. We begin by returning to a fundamental principle, the *conservation* of charge (see Part I, Electrostatics).

Consider a volume in space bounded by the closed surface S. If a net current passes through that surface, either into or out from the volume, then the net charge existing in the volume must change. And the magnitude of that change must be just equal to the magnitude of the net charge carried across that surface. This is simply another way of saying that we cannot create or annihilate net charge ; we can just move charges around. And we can keep track of them.

The mathematical statement of the conservation of charge principle is

$$\mu_0 \sum_{S \text{ closed}} \vec{j} \cdot \Delta\vec{S} = -\frac{\Delta Q}{\Delta t} \tag{9.6}$$

where we have expressed the principle in terms of the time rate of change of charge instead of just the change of charge. Here $\vec{j}$ is the current density on the surface S, and Q is the net charge in that volume enclosed by S. We need the minus sign in equation (9.6), because I is considered positive when the net conventional current is outward through the bounding surface.

For steady currents and constant—charge densities the quantity $\Delta Q/\Delta t$ is zero for any volume, for whatever currents pass into a volume must pass outward, too. If a current passed into a volume without leaving, then positive charge would accumulate in that volume. In that case, the charge density would be changing at some place in the volume, and that is contrary to the assumption that charge densities are constant; i.e., time-independent. Then the mathematical statement of the principle, or law, of charge conservation for the special case when everything is independent of time is just

$$\mu_0 \sum_{S \text{ closed}} \vec{j} \cdot \Delta\vec{S} = 0 \tag{9.7}$$

Now we must be careful. Equation (9.7) is *not* the right-hand side of Ampére's law given by combining equations (9.1) and (9.2),

$$\sum_{s \text{ closed}} \vec{B} \cdot \Delta\vec{s} = \mu_0 \sum_{S} \vec{j} \cdot \Delta\vec{S} \tag{9.8}$$

In equation (9.7) the surface S is *closed*, and it completely bounds a volume. In equation (9.8) the surface is, in general, *not closed*, but it is itself bounded by the closed path s. Nevertheless, charge conservation as given by equation (9.7) *is* contained in Ampére's law as given by equation (9.8). And that is just what assures us that Ampére's law can- not possibly work in time-dependent situations.

To support this last assertion we want to show that the special case of charge conservation, (9.7), is contained in Ampére's law, (9.8). Consider a closed surface S which in our mind's eye we imagine to be cut into the two surfaces S_1 and S_2, as shown in Figure 9.3.

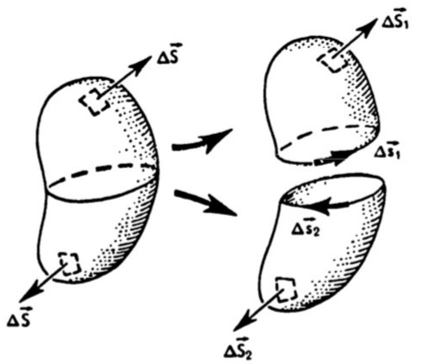


Figure 9.3: Breaking a closed surface into two surfaces, each of which is bounded by a closed path.

We can certainly say that the current passing out from the closed surface S is just that passing "out" through S_1 plus that passing "out" through S_1. (Why are the two "outs" in quotes?) We put this in mathematical form:

$$\sum_{S\ \text{closed}} \vec{j}\cdot\Delta\vec{S} = \sum_{S_1}\vec{j}\cdot\Delta\vec{S}_1 + \sum_{S_2}\vec{j}\cdot\Delta\vec{S}_2$$

There is not really any physics in this mathematical statement, but now we can apply Ampére's law to each of the terms on the right-hand side. If we do put physics into the equation this way, we get

$$\sum_{S\ \text{closed}} \vec{j}\cdot\Delta\vec{S} = \frac{1}{\mu_0}\left(\sum_{s_1\ \text{closed}} \vec{B}\cdot\Delta\vec{s}_1 + \sum_{s_2\ \text{closed}} \vec{B}\cdot\Delta\vec{s}_2\right)$$

To be sure that we are calculating the currents "out" from S_1 and S_2, we must make sure that we traverse the paths s_1 and s_2 in the right senses. You should check to see that $\Delta\vec{s}_1$ and $\Delta\vec{s}_2$ are correctly shown in Figure 9.3.

But s_1 and s_2 are really just the same path. In the first circulation of $\vec{B}$ we are traversing that path in one sense, and in the second we are traversing the same path in the opposite sense. The second circulation then just cancels the first, for surely

$$\sum_{s_1\ \text{closed}} \vec{B}\cdot\Delta\vec{s}_1 = -\sum_{s_2\ \text{closed}} \vec{B}\cdot\Delta\vec{s}_2$$

With this we are left with the result that

$$\sum_{S\ \text{closed}} \vec{j}\cdot\Delta\vec{S} = 0$$

if Ampére's law is valid. And this is just equation (9.7), the conservation of charge principle for the special case of time-independent currents and charge densities

Well, then, we have proved what we set out to prove: namely, Ampére's law contains within it the statement of *time-independent charge conservation*. No wonder it doesn't work for time-dependent cases! In particular, we see why it led to a contradiction in the previous section when we tried to apply it to the charging capacitor: Equation (9.7) simply is not true for any surface that encloses a single plate during charge or discharge.

Now that we know at least one good reason that Ampére's circulation law works only for time-independent situations, we go back to the question of what we can do to generalize it – if we can do anything

9.4 Modified Circulation Law, Displacement Current

Good experimenters do not simply rummage about in a random way, hoping by chance to fall upon new information about the way nature behaves. Faraday, for instance, had something in mind when he began his *Researches*. His early vague thoughts were more hunches than ideas, but he followed them doggedly.

In Chapter 8, we traced Faraday's progress from his first halting steps through the inspired series of experiments that searched out nature's behavior, and finally to the time when he could formulate a new fundamental physical law. The process was: *hunch, experiment, theory*. That is certainly an oversimplification, but it describes in a rough way a process that has often led to new knowledge. But that is not the only process in the scientific enterprise

Sometimes the order is turned around: *hunch, theory, confirming experiment*. That such an order often leads to advances in scientific under-standing is a surprise to some people. They are usually the ones who think that "the scientific method" is the gathering of data until that data forces the recognition of an important order, or law. What we discuss next should serve to bury that narrow conception of what "the scientific method" is .

We are going to follow Maxwell's reasoning as we try to generalize Ampére's law so that it will be valid in time-dependent situations. We shall introduce no new experimental evidence. We shall not really deduce anything. We are going to use our imagina-

tion as we try to "fix up" the law, being careful to avoid introducing relationships we already know to be wrong.

Let's review the state of our understanding. We know that Ampére's law is valid for steady currents and constant-charge densities; i.e., for time-independent situations. Further, we know that Ampére's law contains the charge-conservation principle for time-independent situations, and that alone is enough to assure us that Ampére's law cannot be applicable when there are changing currents or charge densities present.

We might ask ourselves the following question: Since Ampére's law contains the special case of charge conservation, can it be that Ampére's law is itself just a special case of a general circulation law that contains the general principle of charge conservation?

If the arrow below means "implies that," then we can write that

$$\begin{gathered}\sum_s \vec{B}\cdot\Delta\vec{s} = \mu_0 \sum_S \vec{j}\cdot\Delta\vec{S} \\ \Downarrow \\ \sum_S \vec{j}\cdot\Delta\vec{S} = 0\end{gathered} \tag{9.9}$$

for time-independent cases. For our sought-after general circulation law, we might write that

$$\begin{gathered}\sum_{s \text{ closed}} \vec{B}\cdot\Delta\vec{s} = \, ? \\ \Downarrow \\ \sum_S \vec{j}\cdot\Delta\vec{S} + \frac{\Delta Q}{\Delta t} = 0\end{gathered} \tag{9.10}$$

We will have made some progress, if for the ? in equation (2.10) we can put the flux of something through a surface bounded by the circulation path. That would be nice, because if we can write the general circulation law as

$$\sum_{s \text{ closed}} \vec{B}\cdot\Delta\vec{s} = \sum_S \vec{C}\cdot\Delta\vec{S} \tag{9.11}$$

then, even though we don't yet know what $\vec{C}$ is, we are certain that

$$\sum_{S \text{ closed}} \vec{C}\cdot\Delta\vec{S} = 0 \tag{9.12}$$

That (9.12) follows from (9.11) can be shown by exactly the same argument we used to show that (9.7) follows from (9.8); that is, the same argument that let us put the "implies that" arrow in (9.9). That is a purely mathematical argument; there is no physics in it.

Now if we can choose $\vec{C}$ such that (9.12) is the general statement of charge conservation, then we will indeed be making progress. Namely, the general circulation law will imply the general statement of charge conservation. The trouble is that the general statement of charge conservation; i.e.,

$$\sum_{S \text{ closed}} \vec{j} \cdot \Delta\vec{S} + \frac{\Delta Q}{\Delta t} = 0$$

has not come to us in the form of (9.12). The first term is in the right form; it is a flux through a closed surface. But the second term is not. Can we cast that term in the form of a flux of something through a closed surface? We need a relationship between the net charge contained in a volume and the flux of something through the surface enclosing that volume. We do know of such a relationship: Gauss's law (see Part I, Electrostatics).

Gauss's law says that the flux of the electric field intensity $\vec{E}$ through a closed surface is proportional to the net charge encompassed by that surface. In mathematical terms,

$$\sum_{S \text{ closed}} \vec{E} \cdot \Delta\vec{S} = \frac{Q}{\epsilon_0} \tag{9.13}$$

where $\vec{E}$ is the electric-field intensity on the closed surface S, and Q is the net charge in the volume enclosed by S. Using Gauss's law, we get that

$$\frac{\Delta Q}{\Delta t} = \epsilon_0 \frac{\Delta}{\Delta t}\left(\sum_{S \text{ closed}} \vec{E} \cdot \Delta\vec{S}\right)$$

or, if the surface S is held fixed in space,

$$\frac{\Delta Q}{\Delta t} = \sum_{S \text{ closed}} \epsilon_0 \frac{\Delta\vec{E}}{\Delta t} \cdot \Delta\vec{S} \tag{9.14}$$

If we make this substitution, we can write the general conservation

as

$$\begin{aligned}\frac{\Delta Q}{\Delta t} \sum_{S\text{ closed}} \vec{j} \cdot \Delta\vec{S} + \sum_{S\text{ closed}} \epsilon_0 \frac{\Delta\vec{E}}{\Delta t} \cdot \Delta\vec{S} \\ = \sum_{S\text{ closed}} \left(\vec{j} + \epsilon_0 \frac{\Delta\vec{E}}{\Delta t}\right) \cdot \Delta\vec{S} \\ = 0\end{aligned} \tag{9.15}$$

We have been trying to get the general principle of charge conservation into the form of (9.12), and we have done it. The term inside the parentheses in (9.15) is just the $\vec{C}$ we have been after. Then following our lead in (9.11), we write the proposed general circulation law as

$$\sum_{S\text{ closed}} \vec{B} \cdot \Delta\vec{s} = \mu_0 \sum_{S} \left(\vec{j} + \epsilon_0 \frac{\Delta\vec{E}}{\Delta t}\right) \cdot \Delta\vec{S} \tag{9.16}$$

We have put a μ_0 in front of the right-hand side of (9.16), so that the general law will satisfy another condition that we certainly must impose. We want the general circulation law to reduce to Ampére's law for time-independent situations, since Ampére's law is correct in these cases. And (9.16) now does just that; $\Delta\vec{E}/\Delta t$ being zero if everything is independent of time.

Well, we have fixed up the circulation law in such a way that it has some nice features:

(a) For time-independent situations; i.e., steady currents and constant-charge densities, we get Ampére's law back again.

(b) The law has the general, as well as the special, statement of charge conservation built into it.

(c) In a case such as the charging capacitor, Figure 9.2, the right-hand side of (9.16) is not zero, even for a surface that passes between the plates. And we get the same result for that right-hand side no matter what surface we choose, so long as the surface is bounded by the circulation path.

Exercise

Can you prove that the second sentence in (c) above is true? If you cannot, you probably do not understand the conservation of charge principle as it is given by (9.6) or by (9.10).

In developing the general circulation law, we were concerned about certain criteria we knew must be satisfied, if such a law were

to exist at all. And we built in the general conservation principle too. Once begun, everything went along surprisingly well. Something like knocking over the first in a line of toy soldiers. It is encouraging that all went so smoothly, but the fact is that we have been playing a game. Aside from the requirement that we not do violence to anything already known to be true, we set our own rules for the game. We still need to answer the important question: Is our proposed general circulation law true; i.e., does it check with experiment? This is the ultimate question asked of all theory, no matter how pretty that theory may seem. The wonderful thing about what we have done here is that the general law *is* true. It *does* describe the way nature works.

It is not at all clear why our procedure did lead to a physically valid result. We set out to save Ampére's law, but we had no truly guiding experimental results to lead us. Nor did Maxwell. Equation (9.16) is a statement about the physical world. It says that a certain arrangement of measurable things is invariably equal to another arrangement of some other independently measurable things. We did not deduce this relationship from known principles, and so we have no guarantee that the relationship is true. Nevertheless, it is true. But the experimental verification came after Maxwell had proposed it.

We have, of course, presented this development in an artificial way, not at all as its creator likely did it for the first time. Textbook writers are supposed to know where they are going, so we went right on a beeline from the posing of the question to the statement of the proposed answer. And we made nary a wrong turn along the way. Our development comes closer to what might appear in a scientific journal, and that sort of thing disguises or hides all questions, mental gymnastics, and false starts. Nobody reports in a journal how many reams of paper he threw away after running down hunches that did not work out. The report that goes into a journal is usually cleaned up, so that it appears logical and straightforward.

The way we went about getting the general circulation law is a distortion in another sense. Maxwell had in mind much more than just Ampére's law, although that law was vitally important to his entire scheme. He was trying to construct what we today call a *field* theory for all of electromagnetism. He was using the conceptualizations of Faraday, putting them in manageable mathematical form, extending and generalizing the known physical laws, and putting it all together into a concise and consistent whole. And the whole was indeed more than the sum of its parts. We shall see Maxwell's

full theory all in one place in Chapter 11.

In Chapter 10 we shall see how Maxwell's extension of Ampére's law is verified. We shall see that the term $\epsilon_0 \, \Delta\vec{E}/\Delta t$ that we added to the regular conduction current density $\vec{j}$ turns out to be necessary for the description of electromagnetic fields. Maxwell called that added term the *displacement current*. That is probably not very appropriate today, but its root lies in Maxwell's model around which he built much of his theory. The model has lost its cogency, but the label continues.

Chapter 10

Propagation Of An Electromagnetic Disturbance

10.1 The Circulation Laws In Empty Space

We now have two circulation laws. The first, Faraday's law, tells us that the *emf* around any closed path is equal to the negative of the time rate of change of the magnetic flux through any surface bounded by that path.

$$\begin{aligned} \mathcal{E} &\equiv \sum_{s \text{ closed}} \frac{vecF}{q} \cdot \Delta s = -\frac{\Delta \Phi_B}{\Delta t} \\ &\equiv -\frac{\Delta}{\Delta t} \sum_{S} \vec{B} \cdot \Delta \vec{S} \end{aligned} \tag{10.1}$$

The second, Ampére's law as modified by Maxwell, tells us that the circulation of $\vec{B}$ around any closed path is proportional to the sum of two fluxes through any surface bounded by that path: the conduction-current density and the displacement-current density.

$$\sum_{s \text{ closed}} \vec{B} \cdot \Delta \vec{s} = \mu_0 \sum_{S} \left(\vec{j} + \epsilon_0 \frac{\Delta \vec{E}}{\Delta t} \right) \cdot \Delta \vec{S} \tag{10.2}$$

As written here, (10.2) already assumes that the circulation path s is being held fixed in space. That is not the case for (10.1), but we shall be dealing only with those circulation paths which are stationary. Further, (10.2) is written for a circulation path that is in empty space, i.e., everywhere outside material media. That is not true for equation (10.1), for which the circulation path can be within a material loop. But equation (10.1) is certainly valid

when the circulation path is entirely in empty space, and we shall be considering only such paths.

Suppose now that we consider circulation paths which are stationary in empty space, and suppose further that in the region we are investigating there are no charge densities and no conduction currents. If a circulation path is stationary, then $\vec{F}/q = \vec{E}$ in (10.1), and we can write

$$\sum_{s\ \text{closed}} \vec{E} \cdot \Delta\vec{s} = -\frac{\Delta\Phi_B}{\Delta t} = -\frac{\Delta}{\Delta t}\sum_{S} \vec{B} \cdot \Delta\vec{S} \tag{10.3}$$

and if there are no conduction currents around, then $\vec{j} = 0$ in equation (10.2), and we can write

$$\begin{aligned}\sum_{s\ \text{closed}} \vec{B} \cdot \Delta\vec{s} &\equiv \mu_0\epsilon_0 \frac{\Delta\Phi_E}{\Delta t} \\ &= -\mu_0\epsilon_0 \frac{\Delta}{\Delta t}\sum_{S} \vec{E} \cdot \Delta\vec{S}\end{aligned} \tag{10.4}$$

Equations (10.3) and (10.4) show the intimate relationships that exist between electric and magnetic fields in time-dependent circumstances. In such cases, we no longer have two separate areas of interest: electric fields and magnetic fields. When things are changing it is difficult and often misleading to think of the two fields independently. Faraday's law tells us that there is always an electric field associated with a time-dependent magnetic field, even in the absence of charges.

The Maxwell-Ampére law tells us that there is always a magnetic field associated with a time-dependent electric field, even in the absence of conduction currents. Equations (10.3) and (10.4) tell us the interdependence of E and B and allow us to keep them both in mind at once. When we do that, we say that we are talking about an *electromagnetic* field, the description of which requires the descriptions of both $\vec{E}$ and $\vec{B}$.

In Chapter 8, when we discussed Faraday's law, we sometimes said that a changing magnetic flux *creates* or *brings into being* an electric field. In Chapter 9, we thought of a changing electric field as being a *cause* of a magnetic field. As we now look at equations (10.3) and (10.4), we see that this kind of thinking does not make sense. The laws do not tell us which causes which, they only tell us that in time-dependent situations we got both an $\vec{E}$ and a $\vec{B}$. We know only that there are two associated field vectors: $\vec{E}$ and $\vec{B}$. It

makes no sense to say that a changing $\vec{B}$ creates an $\vec{E}$, which if it changes creates a $\vec{B}$. We would need a time sequence of events, if we were to think in this way. And the laws provide no such sequence.

Nor do equations (10.3) and (10.4) tell us about a mechanism through which the two fields sense each other's existence. This seeming deficiency has caused conceptual difficulties for very able physicists. We shall discuss that problem later on when it is more appropriate. First we are going to investigate a very remarkable consequence of the interdependence displayed in equations (10.3) and (10.4).

10.2 The Velocity Of Propagation

Maxwell came to a startling conclusion when he combined the physics contained in our equations (10.3) and (10.4). He convinced himself that light is an electromagnetic disturbance – that light consists of a varying electric field and an associated magnetic field directed perpendicular to each other and both perpendicular to the direction in which light travels.

We are going to develop the evidence that was most convincing to Maxwell, although we shall not do it in quite the way he did. We want to show that a disturbance in an electromagnetic field travels in empty space with the speed of light. We are going to use the two circulation laws as written for empty space.

We direct our attention to an empty region of space. There are no material objects there and thus no charges and no currents. We imagine that there is a changing field somewhere in that region, and we begin with a particular kind, suiting our purpose. Suppose we have a magnetic field such that $\vec{B}$, pointing in the $+z$ direction, is uniform everywhere on one side of a very large plane surface parallel to the yz plane and that $\vec{B}$ is zero everywhere on the other side of the surface. We introduce the changing character of the field by letting this plane travel in the $+x$ direction with the speed v. Then, at every instant, the magnetic field is changing at every point on the moving plane. Figure 10.1 shows this arrangement.

We are assuming that this disturbance is traveling along unabated. We do not want to say anything about how it was produced. At this point in our study we have not yet convinced ourselves that such a disturbance can be produced. We just imagine that the disturbance, as described, exists and that we have no knowledge of its distant history. We know what it is doing now, and we assume that it will

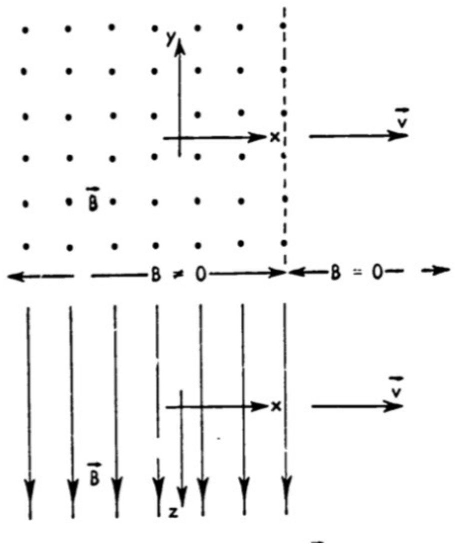


Figure 10.1: The uniform field $\vec{B}$ is to the left of the moving plane and points in the direction of increasing z.

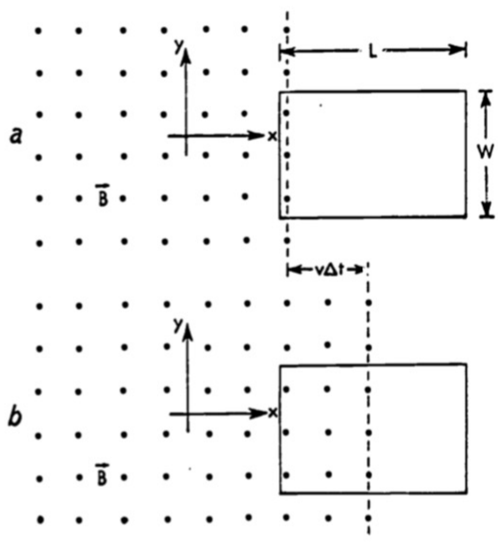


Figure 10.2: (a) The location of the moving plane at the time t. (b) The location of the moving plane at the time $t + \Delta t$.

continue in the same way. And we ask what the characteristics of such a disturbance would, be.

We first use the circulation law given by equation (3.3). We select the stationary rectangular circulation path in the xy plane as shown in Figure 10.2. The left side of the rectangle is inside the region where $\vec{B} \neq 0$, i.e., to the left of the moving plane. Figure 10.2 (a) shows the location of the moving plane at the time t, and Figure 10.2 (b) shows it at the time $t + \Delta t$. We have made certain that the rectangle's length L is large enough, so that the moving plane will not have reached the right side of the rectangle by the time $t + \Delta t$.

We can calculate the change in the the magnetic flux through a surface bounded by this rectangular path. In the time interval Δt the change in that; flux is

$$\Delta\Phi_B = Bwv\,\Delta t,$$

so that the time rate of change of the magnetic flux is

$$\frac{\Delta\Phi_B}{\Delta t} = Bwv \tag{10.5}$$

We shall call this quantity positive when the flux is increasing in time in the $+z$ direction, i.e., when its increase is as shown in Figure 10.2.

Now Faraday's law, equation (10.3), says that the circulation of $\vec{E}$ around that same rectangular path must equal the negative of $\Delta\vec{B}/\Delta t$. Negative, that i.e, when we traverse the rectangle in the counterclockwise sense as seen in Figure 10.2. Then there must be an electric field intensity $\vec{E}$ somewhere. Where is it, what is its magnitude, and in what direction is it pointing?

As we calculate the circulation of $\vec{E}$ around the rectangular path, we go along the lower side in one direction parallel to the x-axis and along the upper side in the opposite direction. These two contributions surely cancel each other, since the field $\vec{E}$ at each point on one of these sides must be equal to $\vec{E}$ at the corresponding point on the other. And we can put the right side of the rectangle so far away from the moving plane that we are certain that no fields exist there, and, thus, there can be no contribution to the circulation along that side. Then only the left side remains to provide a contribution to the circulation of $\vec{E}$, and it must carry the full burden.

The contribution to the circulation along this left-hand side must be $-Ew$, as we see when we recall that we are going in a

counterclockwise sense around the path. That is, for Faraday's law to be correct, $\vec{E}$ must be in the $+y$ direction at all points on the left side of the rectangular circulation path (check this with Lenz's law). Equating the circulation of $\vec{E}$, i.e., $-Ew$, with the negative of equation (10.5), we get for the magnitude of $\vec{E}$ along the left side of the rectangle

$$E = v\,b \tag{10.6}$$

We have assumed that the only component of $\vec{E}$ along the left side of the rectangle is the y component. That is the only component that contributes to the circulation. But when we are finished with this development, it will be clear that $\vec{E}$ has only a y component there.

We can move the left side of the rectangular circulation path anywhere to the left of the moving plane without changing the argument that leads to (10.6). Then there must be an electric field $\vec{E}$ at each point to the left of that moving plane. And the magnitude of that $\vec{E}$ must everywhere be just $v\,B$, and $\vec{E}$ must everywhere point in the $+y$ direction. Then corresponding to the magnetic field shown in Figure 10.1, there must be an electric field as shown in Figure 10.3.

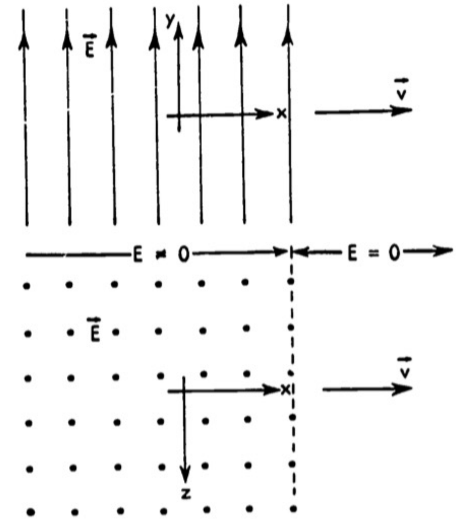


Figure 10.3: (a) The location of the moving plane at the time t. (b) The location of the moving plane at the time $t + \Delta t$.

Exercise

Show that there can be no y component to any electric field intensity that exists at any point to the right of the moving plane.

Of course we can use the circulation law (10.4), too. We now choose the rectangular circulation path in the zx plane as shown in Figure ??, the left side of which is inside the region where $\vec{E} \neq 0$. In the time interval Δt the change in electric flux is

$$\Delta\Phi_E = Ewv\,\Delta t,$$

so that the time rate of change of the magnetic flux is

$$\frac{\Delta\Phi_E}{\Delta t} = Ewv \tag{10.7}$$

We shall call this quantity positive when the flux is increasing in time in the $+y$ direction, i.e., when its increase is as shown in Figure ??.

According to (10.4), the circulation of $\vec{E}$ around the rectangular path should equal $\mu_0\epsilon_0$ times $\Delta\Phi_E/\Delta t$, when we traverse the path

in a counter-clockwise sense as seen in Figure ??. As before, we get a contribution to the circulation only along the left side of the rectangle. If we let $\vec{B}_i$ be the magnetic induction field there, then the circulation gives us just $B_i w$ where B_i points in the $+z$ direction.

(Check this direction using the convention relating the sense of the circulation and the direction of positive flux.) Equating this circulation to $\mu_0\epsilon_0$ times (10.7), we get that

$$E = \frac{1}{\mu_0\epsilon_0 v} = B_i \tag{10.8}$$

Let's review what we have found up to this point. We used Faraday's circulation law on the originally given $\vec{B}$ field, and we found an induced electric field $\vec{E}$ that accompanies $\vec{B}$ everywhere behind the advancing plane. Further, $\vec{E}$ and $\vec{B}$ are perpendicular to each other in such directions that $\vec{E} \times \vec{B}$ is in the direction of the velocity v. Equation (10.6) gives the relations between the magnitudes of $\vec{E}$ and $\vec{B}$.

Next, we used the Maxwell-Ampére circulation law on the electric field $\vec{E}$, and we found that an induced magnetic induction field accompanies $\vec{E}$ everywhere behind the advancing plane. Further, $\vec{B}_i$ is in the same direction as the original $\vec{B}_i$, so that $\vec{E} \times \vec{B}_i$ is in the direction of the velocity v. Equation (10.8) gives the relation between the magnitudes of $\vec{E}$ and $\vec{B}_i$.

To see what this means, we must remember an important point about the circulation laws, i.e., about equation (10.3) and (10.4). If we have time-varying electric and magnetic fields, then the field vectors $\vec{E}$ and $\vec{B}$ must satisfy both equations. Then the $\vec{B}_i$ we have been talking about in connection with the Ampére-Maxwell law cannot be some new magnetic field. It must be the same one we had when we used Faraday's law. That is, vectors $\vec{E}$ and $\vec{B}$ must satisfy equations (10.3) and (10.4) simultaneously. Not only is the $\vec{B}$ in equation (10.3) equal to the $\vec{B}$ in equation (10.4), the $\vec{B}$ in equation (10.3) *is* the $\vec{B}$ in equation (10.4).[5]

How can we build in the fact that $\vec{B}_i = \vec{B}$? Well, $\vec{B}_i$ and $\vec{B}$ are in the same direction, so we just need to arrange things so that their magnitudes are the same. That means that the $\vec{B}_i$ in equation (10.8) must be made to equal the $\vec{B}$ in equation (10.6), which is possible only if

$$v^2 = \frac{1}{\mu_0\epsilon_0} \tag{10.9}$$

[5] In fact, we could have begun with the electric field $\vec{E}$ shown in Figure 10.3 instead of with the magnetic field. We would have used the Maxwell-Ampére circulation law on that $\vec{E}$ field, and *then* we would have used Faraday's circulation law on the associated magnetic field. We would have ended with the same relationships.

That seems strange. The advancing plane cannot have just any velocity, it must have a *particular velocity*. We can find out what that v is by using the known values for μ_0 and ϵ_0 and (see Parts I and II, Electrostatics and Magnetostatics). You may remember them in the forms

$$\frac{1}{4\pi\epsilon_0} = 9 \times 10^{10}\ \mathrm{N\,m^2/C^2}$$
$$\frac{\mu_0}{4\pi} = 9 \times 10^{-7}\ \mathrm{N/A^2}$$

A little rearranging gives the result $1/\mu_0\epsilon_0 = 9 \times 10^{16}\ \mathrm{m^2/s^2}$ or

$$v = \frac{1}{\sqrt{\mu_0\epsilon_0}} = 3 \times 10^8\ \mathrm{m/s} \tag{10.10}$$

But this is not just an interesting velocity, it is the *velocity of light!* [6] Who can imagine the thoughts flashing through Maxwell's mind when this came to him?

[6] In vacuum, which is fine since ϵ_0 and μ_0 are supposed to be determined for vacuum, too. Actually, $c \cong 2.998 \times 10^8$ m/s.

We should be careful about what we have and have not done. We have shown that if there is a self-sustaining electromagnetic disturbance that travels through empty space, then the velocity of propagation of that disturbance is c, the velocity of light in empty space. We have not explained how to create such a disturbance. We have not shown how the disturbance is related to sources, i.e., what the sources must do to create this kind of disturbance.

This is the evidence Maxwell considered most convincing. He reported his conclusion in confident, though guarded, language,

> "This velocity is so nearly that of light, that it seems we have strong reason to conclude that light itself (including radiant heat, and other radiations, if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws."[7]

[7] Maxwell's paper, "A Dynamical Theory of the Electromagnetic Field," *Philosophical Transactions* , Vol. 155, 1865. The quote is taken from *A Source Book in Physics* by Willlam Francis Magie, McGraw-Hill Book Co., New York, 1935

Today there is no doubt; light is electromagnetic in character. And by "light" we mean the entire electromagnetic spectrum: radio waves, microwaves, infrared radiation, visible light, ultraviolet light, rays and y rays. The evidence is now overwhelming. Each of these is an electromagnetic disturbance, the frequency increasing in the order in which they are listed. But conclusive experimental evidence did not come until more than twenty years after Maxwell's work (see Section 10.3).

We chose a particular kind of disturbance for our development, but the conclusions are really independent of that choice. We have taken an extreme case in which the field goes from zero to a finite

value abruptly at the moving plane front, Figure 3.5 (a). We could have chosen to build up the field in space at any instant by a set of smaller steps, Figure 3.5 (b). These steps need not be very widely separated in space, because we can always choose the length L of the rectangular circulation paths to be shorter still. Of course, the corresponding At would need to be smaller, too. In any case, we could treat each boundary between successive steps just as we treated the single one before.

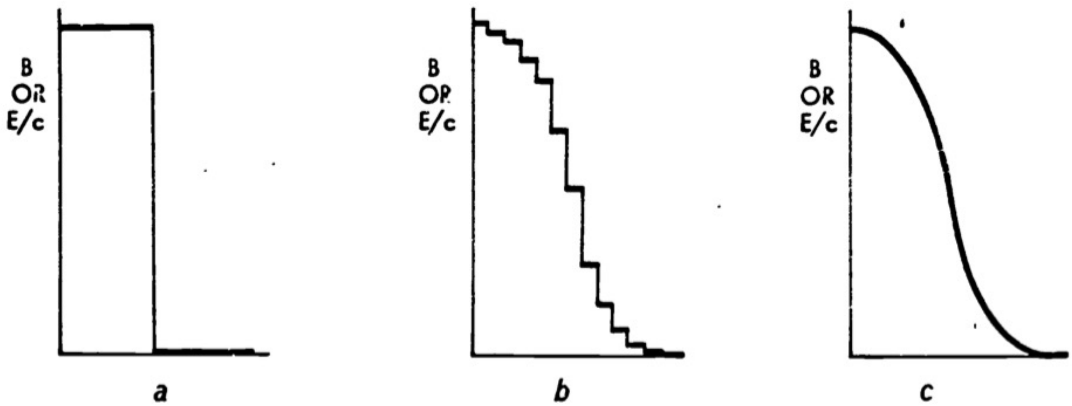


Figure 10.4: Spatial profiles of electromagnetic disturbances: (a) a single abrupt, discontinuous rise; (b)a series of smaller discontinuous rises; and (c) a continuous rise.

And while we are not yet prepared to give a rigorous proof, it is true that the results are the same when the field has any continuously changing profile in space, Figure 3.5 (c).

That is, no matter what the shape of the spatial profile of the electromagnetic disturbance, at every point $E = cB$, and the disturbance travels through empty space while maintaining its shape (see Figure ??. All the traveling disturbances we have described are called *plane waves*, each point in the disturbance traveling in the same direction. We shall restrict ourselves to plane waves in this monograph, but you can likely figure out for yourself some of the characteristics of cylindrical and spherical waves.

We are now in position to see the consequences of what Maxwell did to Ampére's circulation law (see Chapter 9). He not only saved the law for time-dependent situation , i.e. predicted the character of electromagnetic radiation. The displacement current $\mu_0\epsilon_0(\Delta\vec{E}/\Delta t)$ that he added, see equations equations (10.2) and (10.4), is just what makes that prediction possible. This is surely the strongest evidence we have for Maxwell's modification of Ampére's law.

Further, we begin to see how the entire theory fits together. We

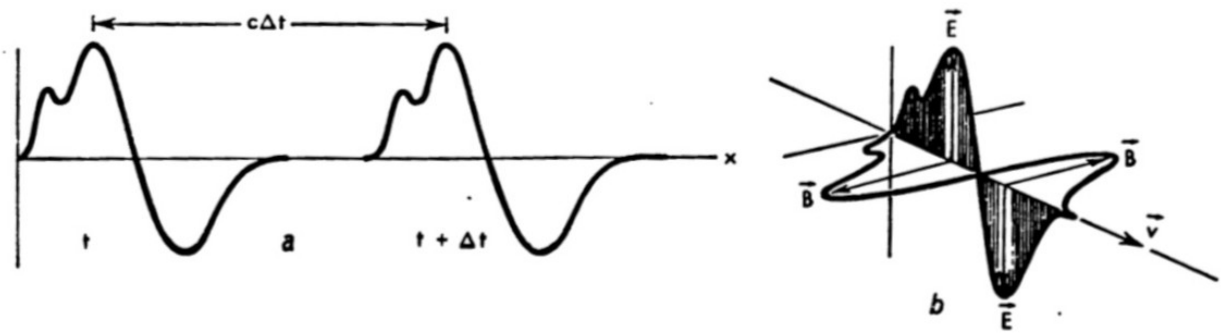


Figure 10.5: The characteristics of a plane electromagnetic wave in empty apace. (a) The spatial profile is unchanged as it travels with the speed c. (b) $\vec{E}$ and $\vec{B}$ are perpendicular, and at each point $\vec{E} \times \vec{B}$ is in the direction of the wave's velocity. At each point $E = c\,B$.

now have four independent statements about electromagnetic fields in empty space:

$$\sum_{S \text{ closed}} \vec{E} \cdot \Delta\vec{S} = \frac{Q}{\epsilon_0} \qquad \text{(a)}$$

$$\sum_{S \text{ closed}} \vec{B} \cdot \Delta\vec{S} = 0 \qquad \text{(b)}$$

$$\sum_{s \text{ closed}} \vec{E} \cdot \Delta\vec{S} = -\sum_{S} \frac{\Delta\vec{B}}{\Delta t} \cdot \Delta\vec{S} \qquad \text{(c)}$$

$$\sum_{s \text{ closed}} \vec{B} \cdot \Delta\vec{s} = \mu_0 \sum_{S} \left(\vec{j} + \epsilon_0 \frac{\Delta\vec{E}}{\Delta t}\right) \qquad \text{(d)}$$

$$(10.11)$$

where the circulation paths are assumed to be at rest in the frame in which $\vec{E}$ and $\vec{B}$ are measured. These four equations along with the Lorentz force law contain all of electromagnetic theory. You should be able to say, in words, what each of these tells us about the way nature behaves. And you should be able to explain the physical basis for each and give an example or two of situations that each describes. Further, you should now see that, taken together, these equations tell us something that they did not when taken one at a time.

These are the famous Maxwell equations, here written for the case when the field point is in empty space. You will meet them again in Chapter 11 [not yet completed]. At that time we shall be able to use a bit more sophisticated mathematics to help us, but we already have all the physical concepts we need.

Before leaving this section, we should make one further point. Starting with the monograph Electrostatics, following with Magnetostatics, and now in this monograph, we have been proceeding as

if the two constants ϵ_0 and μ_0 were independent of each other and were, thus, independently defined and measured. But if we take the point of view, which we now do, that c is a universal constant, then we have just one of these constants to define. The choice is to define μ_0 as being $4\pi \times 10^7$ N/A^2*exactly*. Then ϵ_0 is no longer an independently defined quantity; it is given by $\frac{1}{(c^2\mu_0)}$.